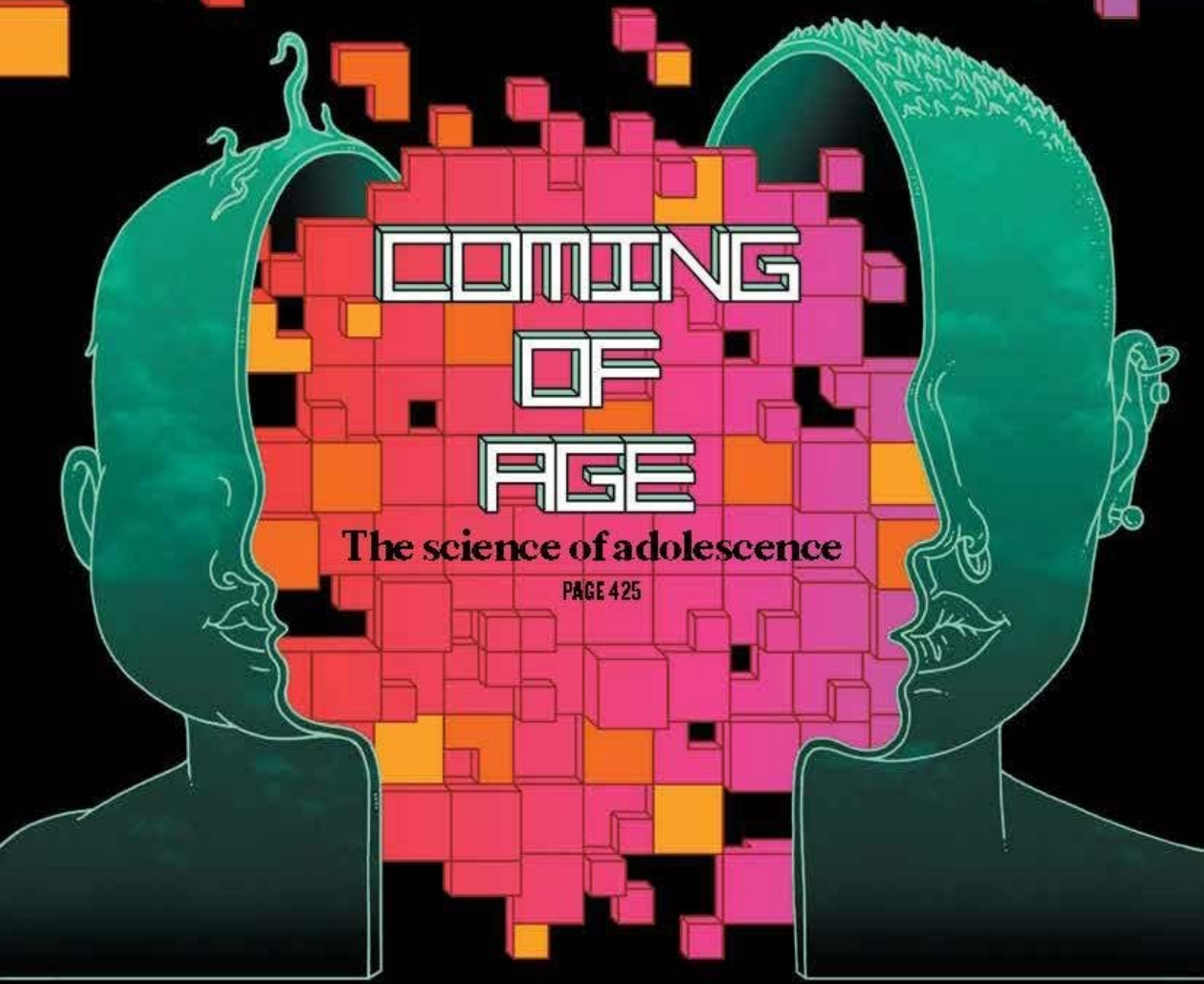
nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



COMING OF AGE

The science of adolescence

PAGE 425

ENGINEERING

MEET THE MEMTRANSISTOR

Multi-terminal device mimics complex neural functions

PAGES 472 & 500

DISEASE

REGULATORY FAILURE

Transcription factor linked to pancreatic inflammation

PAGES 470 & 533

MECHANOCHEMISTRY

UNDER PRESSURE

How compression can force molecules to react

PAGES 468 & 505

NATURE.COM/NATURE

22 February 2018

Vol. 554, No. 7693

THIS WEEK

EDITORIALS

CLIMATE Beware the 'magical thinking' of negative emissions **p.404**

WORLD VIEW Stamp out xenophobia in global science **p.405**



REMAINS Watery grave discovered for prehistoric Swedes **p.407**

Adolescence science must grow up

Young people get a raw deal from society. Targeted study and approaches as part of a new global effort are urgently needed to help them.

The journalist Earl Wilson wrote that snow and adolescence are the only problems that disappear if ignored for long enough. When it comes to science- and evidence-based approaches to welfare, adolescence has been ignored for too long. That needs to change.

This week, *Nature* and the Nature journals try to point the way, to offer support to those living through a too-often misunderstood phase of life. In a special package (see page 425 and https://www.nature.com/collections/adolescence), we examine the complexity and promise of adolescence, and assess problems this age group faces, as well as possible solutions, through the lens of disciplines from medicine and social science to education and neuroscience. The need has never been greater: 10–24-year-olds now make up a record 25% of the global population.

When discussing this generation, it is easy — especially for those in rich societies — to resort to clichés about rebellious attitudes and eccentric behaviour. It's worth remembering that both childhood and adolescence are cut short in many parts of the world, with young people working in hazardous conditions or bearing children in their teens. Millions of adolescents are forced to grow up too fast.

It can be one of the most creative times of life. But adolescence is also a difficult stage for many; not all development during these years is positive, especially given the increasingly digitized and socially competitive environment in which many teenagers live. Mental disorders such as depression and alcohol abuse are particularly likely to strike. And the pile-up of multiple problems, such as poor nutrition compounded by stress, is now understood to have lifelong consequences, including obesity and diabetes.

Too many people in science and medicine share society's indifferent attitude: adolescence is a phase to be endured and moved on from as quickly as possible. That is wrong. A modern health-care system without a focus on the unique challenges of paediatrics or geriatrics would be unthinkable, yet there is no similar effort on behalf of adolescents.

It's not just science. The international-development community has never sufficiently prioritized adolescence. In the past decades, major aid and development organizations and governments have come to agree on the importance of a child's first 1,000 days of development. There has been no such emphasis on what comes later. The United Nations Millennium Development Goals set (important) targets for maternal health, child mortality and primary education, but teenagers were left out.

To improve things, first, the scientific community needs to appreciate what it doesn't know. For infants, there exists a wealth of knowledge on normative growth, large-scale studies on the impacts of early deprivation and a rich evidence base for how to intervene positively in early development. No comparable body of knowledge exists for adolescents. Nobody knows for sure what it means to be a typically developing teenager, nor the best way to improve health or behaviour and to address young people's mental-health conditions. And there are no good ways to measure progress towards improving the lives of adolescents around the world. (For babies, investigators can track factors including mortality, stunting and vaccination rates.)

Next, society needs to be intelligent about interventions and programmes for young people. More-informed and thoughtful approaches are needed to study and address cognitive and social changes that happen during adolescence, and to design effective interventions. The input of adolescents themselves is crucial here: no scientific programme should be launched without talking to the people it aims to reach.

Such programmes shouldn't assume knowledge on the basis of what researchers know about children and adults. They need major efforts to study the development of adolescents, and their thinking and challenges.

"The input of adolescents themselves is crucial."

Take HIV: despite tremendous progress in other age groups, the virus remains one of the main killers and causes of disability in adolescents. And then there's mental illness: fewer than half of the young people who need psychiatric services in the United States receive treatment, and even fewer get effective specialized therapy. The situation in other parts of the world is much worse. Both of these crises are ripe for targeted approaches by agencies, policymakers and funders. Concerted efforts can produce a tangible improvement: between 1990 and 2016, a global focus on infant health halved the mortality rate for children under five years old.

Many of the issues are highly affected by social and cultural factors, and hence regional ones, and need to be tackled on that scale. But scientists could aim, for example, for global sharing of brain-development data.

It's tough growing up. In many cultures, parents, educators, doctors and policymakers criticize adolescents for their impulsivity, misunderstand their rage and mock their language and customs. The next generation deserves better. And we can start by paying attention. ■

Blurred distinction

The idea of research excellence is ubiquitous, but what it means depends on the context.

Excellence is everywhere in science. Or that seems to be the plan: to make excellence ubiquitous in research. This month, the University of the West Indies in Kingston, Jamaica, became the latest academic institution to encourage its scientists to excel, setting up a Regional Centre for Research Excellence in the Caribbean.

To be good is no longer enough — excellence, by definition, must go beyond that.

And for those who achieve it — from individual researchers and

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

laboratories, to universities, regions and even entire countries — grants, students and political patronage follow. Britain's largest biomedical-research funder, the Wellcome Trust in London, runs a grant scheme aimed at "Sustaining Excellence", and the United Kingdom funds universities according to a mammoth Research Excellence Framework (REF) assessment every few years.

What does excellence mean? How is it measured? When do we know that we have reached the required standard? These are difficult questions, but if the excellence agenda is to be taken seriously, they must be asked — even if they cannot be adequately answered.

A paper in *Science and Public Policy* makes the latest attempt to ask — and indeed answer — them (F. Ferretti *et al. Sci. Publ. Pol.* http://doi.org/ckpg; 2018). The authors interview a dozen experts — from policy wonks to researchers — about excellence and quickly reach two points of consensus.

First, the idea of excellence as a measure of research quality makes many people uncomfortable. And second, these people — despite their discomfort — cannot suggest anything better, given that science and scientists must meet political demands of accountability and assessment.

These arguments will be familiar to those who follow the debate, but the conclusions of the study are still striking. The authors suggest that "the making of current indicators for research policy in the EU may be in need of serious review". This is especially noteworthy because it is those very authors who devised the policy indicators — based, of course, on excellence.

The majority of the authors work at the European Commission's Joint Research Centre (JRC) in Ispra, Italy, which in 2013 took the excellence agenda to its logical conclusion and set up a way to assess the scientific performance of nations. Policymakers in Europe now use this metric — the Research Excellence in Science & Technology indicator — to rank the performance of the member states, and so to set priorities and distribute funds.

Critics of the concept of research excellence (and there are many) will welcome the suggestion from the JRC excellence architects in the new paper that the system is flawed. But the scientific community should remember the second point of consensus identified in the study: if not excellence, then what?

Many scientists would like to see excellence metrics — indeed, all metrics — scrapped. Leave the job of directing research, they say, to researchers. Others suggest that the excellence effort should be rebranded to reflect its most important features — such as "soundness" and "capacity" (S. Moore *et al. Palgrave Commun.* http://doi.org/ckph; 2017).

The case for abandoning metrics is not realistic and not desirable: applied properly, metrics can indeed be a useful guide to policymakers and a way for the public to trace the billions of tax dollars funnelled into research every year. (This is especially the case in countries susceptible to cronyism and nepotism.) And to change the language used is politically unwise. Semantics matter — and excellence, to an extent, is what politicians and policymakers expect from scientists.

"Some funders are starting to place more importance on the societal impact and relevance of research."

But it is true that excellence can be defined in many ways. And this is where reforms should focus. *Nature*, for example, intends to promote the health of research groups this year and, with that, the responsibilities of principal investigators and other group leaders to promote reproducibility. Can a university that does not offer adequate training to people in these positions truly be considered excellent?

Meanwhile, some funders are starting to place more importance on the societal impact and relevance of research. Britain's REF exercise, for example, deserves credit for including such impacts in its assessment. And in recent years, the handling of issues such as equity and social justice have come under welcome scrutiny.

Perhaps most important, in both defining and applying excellence, is transparency. Local definitions can create problems. Young scientists trained at universities that downplay the need for high-impact papers, for example, can find themselves at a disadvantage when applying for jobs at places that attach greater value to them.

Excellence depends on context. But scientists, funders and officials can do more to discuss and agree on some suitable basic principles. A news story last week, for example, revealed that more than three-quarters of research organizations in the United Kingdom have no policy for preventing the misuse of metrics in hiring decisions. Many of these universities consider themselves excellent. Others will disagree. ■

Negative thinking

Work on how rocks draw carbon from the air shows the scale of the emissions challenge.

Decarbonization of the world's economy would bring colossal disruption of the status quo. It's a desire to avoid that change — political, financial and otherwise — that drives many of the climate sceptics. Still, as this journal has noted numerous times, it's clear that many policymakers who argue that emissions must be curbed, and fast, don't seem to appreciate the scale of what's required.

According to the Intergovernmental Panel on Climate Change (IPCC), carbon emissions must peak in the next couple of decades and then fall steeply for the world to avoid a 2 °C rise. A peak in emissions seems possible given that the annual rise in carbon pollution stalled between 2014 and 2016, but it's the projected decline that gives climate scientists nightmares.

The 2015 Paris agreement gave politicians an answer: negative emissions. Technology to reduce the amount of carbon already in the atmosphere will buy society valuable time. The agreement went as far as arguing that incorporating one such technology — bioenergy with carbon capture and storage (BECCS) — could even see the global temperature increase kept to 1.5 °C.

What would negative emissions look like? A Perspective this week in *Nature Plants* offers another glimpse, and it's not pretty (D. J. Beerling *et al. Nature Plants* http://dx.doi.org/10.1038/s41477-018-0108-y; 2018). The review focuses on the idea of enhanced weathering, which aims to exploit how many rocks react with carbon dioxide and water to form alkaline solutions that, over time, find their way into the sea. It's one of a number of proposed negative-emissions technologies.

In theory, enhanced weathering could lock up significant amounts of atmospheric carbon in the deep ocean. But the effort required is astounding. The article estimates that grinding up 10–50 tonnes of basalt rock and applying it to each of some 70 million hectares — an area about the size of Texas — of US agricultural land every year would soak up 13% of the annual global emissions from agriculture. That still leaves an awful lot of carbon up there, even after all the quarrying, grinding, transporting and spreading.

It's not hard to see why many climate scientists have dismissed the near-impossible scale of required negative emissions as "magical thinking". Or why the European Academies' Science Advisory Council said in a report this month: "Negative emission technologies may have a useful role to play but, on the basis of current information, not at the levels required to compensate for inadequate mitigation measures."

The IPCC is now working on a report on strategies to keep warming to under 1.5 °C, which is due to be published later this year. By necessity, those strategies will lean heavily on negative emissions. Scientists must continue to spell out to policymakers the harsh reality of what this would involve, and in the strongest possible terms. ■

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Ask me about my science, not your stereotypes

I've encountered even more prejudice as a researcher from the Middle East than as a woman working in Saudi Arabia, says **Malak Abedalthagafi**.

A reporter from a world-famous newspaper contacted me a few months ago to talk about science in Saudi Arabia. I was excited by the foreign outlet's interest in my country, where women are advancing in fields that include aerospace, engineering, finance and medicine. But all she wanted to hear were my thoughts on headscarves and driving.

Yes, all of the common complaints are true. Yes, women will soon be allowed to drive. Yes, we wear scarves. We even need permission to travel. Our society is still antiquated in so many ways.

But there is another reality I yearn to discuss, one that many of my foreign colleagues are blind to. It is this: women within Saudi Arabia have long done, and long been recognized as doing, real science.

My first inkling that I could become a scientist came when I was hospitalized as a young girl in the 1980s. I have a genetic condition, and had baulked at taking my medicine. After I received surgery in London, my parents announced that they would take me to the best genetic specialist in Saudi Arabia. I sat waiting with dozens of children. We seemed so different. Some appeared healthy; others looked very old or had to breathe through holes in their necks.

When the specialist arrived, I was shocked. She led a group of doctors around the clinic; she dictated while they scribbled furiously. She was strong and serious and did not wear a headscarf.

I later learnt that Nadia Awni Sakati had established one of the first genetics departments in Saudi Arabia. It was clear that she had won the respect of her peers and in the lab was treated as equal to any man. Sakati was a fixture in my life growing up. I was lucky that my parents were able to take me to her clinic, a plane ride away, nearly every month.

I wanted to be just like her.

After medical school at King Abdulaziz University in Jeddah, I studied abroad as part of the King Abdullah Scholarship Program, which has now sent more than 200,000 Saudis around the world to study at prestigious universities as undergraduates, graduate students and clinical postgraduates. Although they are required to travel abroad with a guardian, nearly half the recipients have been female.

I worked at Harvard Medical School and affiliated hospitals in Boston, Massachusetts; at Georgetown University in Washington DC; and at the University of California, San Francisco. I knew the United States was a leader in genetics, and I wanted to get the finest education in the world.

Unfortunately, I was made to feel that the stain of being from a developing country was indelible.

Some of my US colleagues refused to believe I could have received an adequate education at home. I was constantly asked whether I could speak English, whether I had learnt this or that basic concept in medical school, or — as a Saudi woman — whether I would have problems working alongside men. It was a shock to hear such questions from people at world-class institutions.

Once, some colleagues on a new project invited me to happy hour. I was delighted to be included, but nervously admitted that I did not drink. A senior member of the group belittled me. "What, are you afraid you'll be beheaded when you go back home?"

Those comments stung. Until then, I had felt like an equal in the lab. Now I was reminded that, in their eyes, I would always be different.

When I heard about the Saudi national genome project in 2014, I knew I had to go home. I wanted to go back to where I felt I belonged, to be close to my family and to help unlock the genome of my people. Generations of intermarriage means that Saudi Arabia holds many rare genetic disorders. Its scientific productivity has expanded swiftly over the past five years. It is now ranked between Ireland and Argentina for high-quality research output in the Nature Index.

I WAS MADE TO FEEL THAT THE **STAIN** OF BEING FROM A **DEVELOPING** COUNTRY WAS **INDELIBLE.**

Within weeks of my arrival, Sakati called to welcome me home. In the lab, I am treated with respect. My accomplishments and my science stand for themselves.

I cannot always count on that elsewhere. At a conference in North America last year, I described a patient we had treated in our hospital in Riyadh, a five-year-old girl with a brain tumour. We had exhausted all established options; her family opted for end-of-life care.

After carefully reviewing her case, confirming her genetic profile, getting the necessary ethical approvals and having a detailed conversation with her parents, I advised a member of my team to treat her with some experimental medication.

An oncologist from North America was furious at my account. "How dare your team treat this patient?" he fumed, assuming we lacked the clinical capacity to deliver new therapies. When he heard the rest of what I had to say about the case, he apologized.

For scientists from Saudi Arabia and other developing countries, it's important to remember that stereotyping is not a conscious thing. We need to be ready to face it. Those from Western countries should work hard to see past their preconceived ideas.

Meanwhile, that little girl has another chance at life. She is continuing to improve, and remembering to take her medication. ■

Malak Abedalthagafi *is chair of genomics research at King Fahad Medical City and a neuropathologist at King Abdulaziz City for Science and Technology, Saudi Arabia.*
e-mail: malthagafi@kacst.edu.sa

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

SEVEN DAYS *The news in brief*

POLITICS

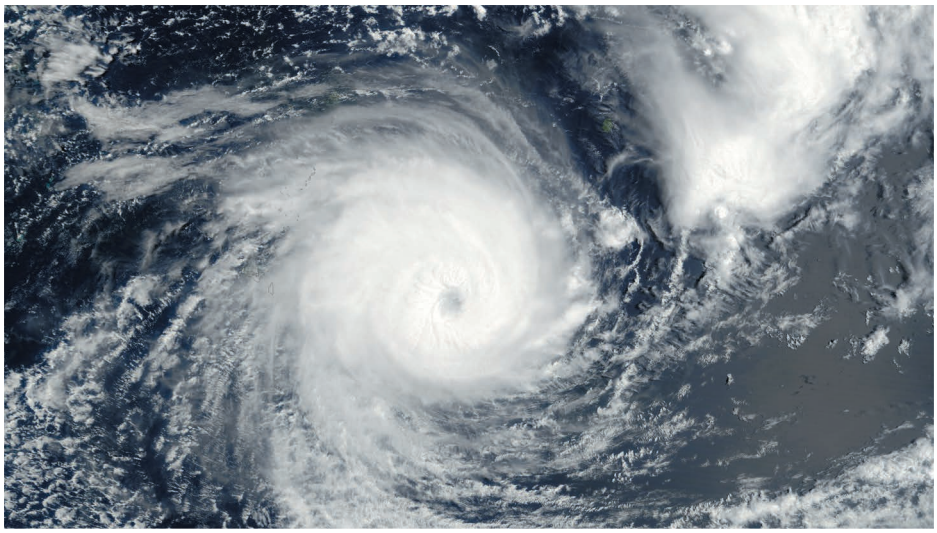
African shifts

Two embattled African leaders resigned last week. South Africa's controversial President Jacob Zuma stepped down on 14 February after pressure from his party, the ruling African National Congress. Zuma's deputy, Cyril Ramaphosa, has replaced him in a move that is expected to bolster international investment confidence in the nation. Ramaphosa's long-time ally Naledi Pandor is South Africa's science minister. The following day, Ethiopian Prime Minister Hailemariam Desalegn resigned, without naming a successor. The East African country, which is the second most populous on the continent and famous for its archaeological sites, has had three years of civil unrest. Desalegn cited political instability as a reason for his sudden resignation and said that he will continue in the role until his successor is chosen at the next congress of the ruling Ethiopian People's Revolutionary Democratic Front.

AWARDS

Wolf prizes

Quantum-communication pioneers Gilles Brassard at the University of Montreal in Canada and Charles Bennett of IBM in Yorktown Heights, New York, were among the recipients of the 2018 Wolf Prizes on 12 February. In 1984, Brassard and Bennett invented the first technique for sharing an encryption key using quantum effects, which would reveal any eavesdropper's attempt to steal the key. In 1993, they also made the first theoretical proposal for using the quantum phenomenon of entanglement — dubbed spooky action at a distance — to teleport quantum states. The two researchers will share the US$100,000 physics prize. The Israeli foundation also awarded $100,000 prizes in agriculture (for honeybee genomics), chemistry (for reticular materials, such as metal–organic frameworks), mathematics and music.



NASA EARTH OBSERVATORY

Cyclone Gita wreaks havoc in Pacific

Tropical Cyclone Gita has ravaged several Pacific island countries. The hardest hit was Tonga, where Gita made landfall on 12 February as a category 4 cyclone — with winds reaching 230 kilometres per hour. The cyclone passed over the most populous island, Tongatapu, destroying more than 100 buildings in the country's capital, Nuku'alofa. Also affected were the nations of Samoa, American Samoa, Fiji and New Caledonia. As *Nature* went to press, Gita had weakened to a category 2 storm, having made landfall in New Zealand on Tuesday.

UNIVERSITIES

UK fees review

UK Prime Minister Theresa May said on 19 February that her government would conduct a review of tuition fees for higher education. In 2012, a previous Conservative-led government controversially tripled annual university fees in England and Wales; they now stand at £9,250 (US$12,960). The review will look at whether courses that cost less to run, such as those in the arts and humanities, should be cheaper than science and engineering courses. Critics say that higher fees could discourage students from taking science degrees and that research departments could suffer. The review will run until early 2019.

PEOPLE

Astronaut selection

Suzanna Randall, an astronomer at the European Southern Observatory (ESO) in Garching, Germany, has been chosen as a finalist in a private initiative to train the first female German astronaut. She will step in to replace a candidate who left the project, called Die Astronautin. Randall will compete against Insa Thiele-Eich, a meteorologist at the University of Bonn and the daughter of European Space Agency astronaut Gerhard Thiele, for a chance to fly on the International Space Station. The ESO announced the pick on 16 February.

EVENTS

Lassa outbreak

The World Health Organization (WHO) is escalating its response to an outbreak of Lassa fever in Nigeria, according to a 13 February announcement. The agency will ship protective and diagnostic equipment to the country and send people to help coordinate treatment and do community outreach. Since 1 January, there have been about 615 suspected cases of the viral haemorrhagic fever in the country, of which 193 have been confirmed. The tally

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

includes 57 deaths for which Lassa was confirmed as the cause. Although the illness is endemic in parts of West Africa, the current number of cases is unusually large.

Mugabe PhD

University of Zimbabwe vice-chancellor Levi Nyagura was arrested on 16 February for his suspected role in awarding an allegedly undeserved PhD to Zimbabwe's former first lady, Grace Mugabe. The wife of toppled leader Robert Mugabe received a doctorate in sociology from the institution in 2014. Staff questioned the degree's credibility at the time, because Mugabe had allegedly completed the work in just three months. The university published the thesis only in January this year after it emerged that Zimbabwe's anti-corruption watchdog was investigating the matter. Nyagura has been released on bail, but is expected to return to court on 5 March, according to local media, which added that Claude Mararike, Mugabe's supervisor, is also facing arrest.

Death campaign

Mathematician and member of the French parliament Cédric Villani has called on the French state to formally recognize the French military's role in the 1957 death of mathematician Maurice Audin (**pictured**, with his



PAUL FEARN/ALAMY

wife) during the Algerian War of Independence. Audin, a French university lecturer who was a member of the Algerian Communist Party, disappeared during the war. The authorities claimed that he had escaped after being arrested in Algiers for harbouring Communist-Party members; he is now widely believed to have been tortured and killed by the French Army. Villani held a press conference with a member of the French Communist Party on 14 February, Audin's birth date, to publicize the case. He told *Nature* that he hopes more witnesses will come forward "to indicate there is hardly any doubt" about the military's role in Audin's death.

POLICY

Species lawsuit

Nine conservation organizations are suing the US government over changes to water regulations, which they say could harm endangered species. On 14 February, the Center for Biological Diversity and other groups filed a formal intent to sue the US Environmental Protection Agency (EPA) and the Army Corps of Engineers for suspending the implementation of a rule that redefines which bodies of water receive federal protection. On 6 February, the EPA and the Army Corps said that they would suspend the rule for two years while they review it. The Southern Environmental Law Center in Charlottesville, Virginia, and 11 states filed separate legal challenges to the suspension decision earlier this month.

Methane rule

On 12 February, the US Bureau of Land Management proposed to revise a rule requiring oil and gas producers to reduce the amount of methane that they burn off or vent to the atmosphere. Methane, a powerful greenhouse gas, is often treated as a waste product by drillers because it's less valuable than oil. The original rule, finalized by the administration of then-president Barack Obama in 2016, was intended to reduce waste and combat climate change. The proposed revisions, pushed for by the oil and gas industry, would effectively eliminate the new waste-reduction requirements. The US government now claims that the 2016 rule imposed excessive costs on industry, and would inhibit US "energy dominance".

Water management

A landmark agreement in Australia for managing water in the Murray–Darling Basin — a major source of the nation's food crops — is on the verge of collapse. In a Senate vote on 14 February, the opposition Green and Labor parties blocked changes to the 2012 agreement that would have allowed farms upstream in the basin's north to use an additional 70 billion litres of water every year, reducing the amount returned to the environment from 390 billion to 320 billion litres. The water minister of the state of New South Wales, Niall Blair, said he would begin withdrawing the state from the plan. The state of Victoria also condemned the vote, with its water minister Lisa Neville saying the entire plan was at risk. An attempt by either state to withdraw from the plan could provoke a legal challenge from the downstream state of South Australia.

Brain-injury test

On 14 February, the US Food and Drug Administration approved the first blood test for concussion. The test, known as the Brain Trauma Indicator, measures the levels of two proteins that the brain releases after suffering an injury. Until now, physicians have had to rely on brain scans to diagnose concussion, but many people with head injuries show no visible signs of trauma. Test manufacturer Banyan Biomarkers in San Diego, California, says that the test is 97.5% accurate in predicting whether a person has brain lesions, which could help physicians decide whether a brain scan is necessary.

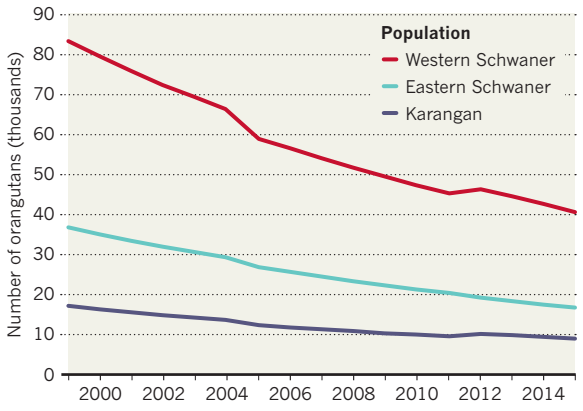
NATURE.COM
For daily news updates see:
www.nature.com/news

TREND WATCH

The island of Borneo lost more than half of its orangutans (*Pongo pygmaeus*) between 1999 and 2015. Researchers estimate that about 148,500 animals died in that time, many as a result of deforestation driven by logging, palm-oil production and mining. Hunting was also a major factor in the animals' decline, say that the team, which compiled data on nest observations to extrapolate the size of the island's orangutan population during the study. The researchers estimate that 70,000–100,000 orangutans are left.

ORANGUTANS IN PERIL

The number of orangutans in Borneo's three largest populations has declined steadily since 1999.



SOURCE: M. VOIGT *ET AL. CURR. BIOL.* HTTP://DOI.ORG/GCX4P2 (2018)

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



DARREN WHITESIDE/REUTERS

Indonesian scientists were hopeful a dedicated fund would reverse years of insufficient investment.

foreign donors, the Indonesian government committed US$3 million to the agency in its first year, through the finance ministry's Indonesia Endowment Fund for Education. The endowment fund also signed a memorandum of understanding with the ISF, agreeing to give the agency money for five years.

Teguh Rahardjo, the ISF's executive director, says that the agency has received some of this cash, but not enough to support all of its grant programmes. Government budget constraints have contributed to the funding shortfall, and the agency is now searching for other sources of money, he says.

He says that the ISF was able to contribute 12 billion rupiah this year towards research grants for collaborations between Indonesian and UK universities. The programme is also supported by the Newton Fund, a British research-development group.

Indonesian researchers have typically had to rely on yearly funding from the national budget, which invests only 0.08% of the country's gross domestic product in scientific research each year. Some scientists argue that the process by which this money is awarded to universities or research centres by various ministries, including the Ministry of Research, Technology and Higher Education, isn't open and transparent. Syarif Hidayat, head of legal, cooperation and information services at the ministry, denies that the funding process is not open and says the ministry is working on ways to stretch research funding over multiple years.

The ISF was established to support long-term research. It is inspired by funding agencies such as the US National Science Foundation, and uses independent peer review to award funding.

"It is very unfortunate that ISF postpones this grant," says molecular biologist Riza Putranto at the Indonesian Research Institute for Biotechnology and Bioindustry in Bogor. "We cannot rely on national budget alone to conduct high-quality research in Indonesia," says Putranto, whose ISF funding application was not successful.

The Ministry of Finance did not respond to *Nature*'s questions about budget constraints. ■

CORRECTION

The Editorial 'Restore justice in Turkey' (*Nature* **554,** 145; 2018) stated that the text-messaging system ByLock was available only through personal introduction. In fact, it was widely available and could have been downloaded by anyone.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

NEWS IN FOCUS

PHYSICS Antimatter experiments seek to probe structure of nuclei **p.412**

OCEANOGRAPHY Sensor network provides clues about Atlantic Ocean current **p.413**

GENOMICS CRISPR-based technique records data from events inside cells **p.414**

NEUROSCIENCE The surprising dexterity with which teens manage risks **p.426**

IVAN ROMANO/GETTY



Luigi Di Maio leads the populist Five Star Movement, which is tipped to receive the highest number of votes in Italy's March election.

ITALY'S ELECTION

Italian science out in the cold

Election campaign gives researchers little hope for future of the nation's science system.

BY ALISON ABBOTT

As campaigning ahead of Italy's national election enters its final weeks, researchers in the country fear that budget cuts and declining interest in science will only continue — whatever the outcome of the vote on 4 March.

A complex coalition government is likely to emerge. The country's traditional centre-left and centre-right parties have splintered, and myriad small parties make up the ballot sheet, as well as the populist Five Star Movement. Topics such as immigration, the refugee influx and eurozone membership have dominated mainstream debates.

But, apart from a battle over the nation's compulsory vaccination programme, which was introduced last year, science has featured little in the campaigning — even as economists warn that Italy's research system is in a precarious state. "We are on the verge of collapse," says Mario Pianta, an economist at the University of Rome Tre, who helps to prepare Italy's statistics on research and development (R&D) for the European Commission.

Italy has hotspots of scientific excellence, such as in particle physics and biomedicine. But, unlike many other European countries, it has failed to modernize its science system in the past few decades. Budgets have constantly been low. Academic hiring practices can be complicated, and bureaucracy crippling, many scientists say. Research organizations have had little power politically, and have been unable to stem the rising influence of those who have demonized vaccinations and promoted charlatan cure-alls.

The gap in scientific achievement and investment between the country's affluent north and poorer south is widening, helping to fuel regionalist and populist politics, says Raffaella Rumiati, vice-president of Italy's national research-evaluation agency, ANVUR. In January, the agency announced the results of its first competition to reward the best-performing university departments, and northern institutions received an overwhelmingly greater share of the funds.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

The outgoing centre-left coalition government, led by Paolo Gentiloni of the Democratic Party, has introduced some research initiatives, including a €1.5-billion (US$1.9-billion) research centre in Milan focused on genomics and personalized medicine, called the Human Technopole. The Democratic Party has some science-related policies in its manifesto that promise more money, research positions and institutional competition.

Pianta says that further reforms to the research system must be supported by increased budgets. But since the 2008 economic crisis, Italy's already low R&D spending has declined by 20% in real terms — equivalent to a hefty €1.2 billion. In 2016, it stood at €8.7 billion. The university budget has shrunk by about one-fifth — to €7 billion — as has the number of professors nationwide. Funding for public research institutes is no higher than it was in 2008, representing a 9% drop in real terms. And Italy's substantial deficit means the situation is unlikely to improve soon.

Even worse, more scientists have left the country since 2008 than have entered it, according to statistics from the Organisation for Economic Co-operation and Development. "It is not just that scientists are going to countries with strong bases in science," says Pianta. "There is also a net loss of scientists from Italy to countries like Spain."

Paradoxically, science is performing well overall. Since 2005, Italy has increased its contribution to the top 10% of the world's most cited scientific documents. And it produces more publications per unit of R&D expenditure than any other European Union country except the United Kingdom. "The happy paradox cannot sustain," says Pianta. "We are heading towards mediocrity."

The next government will have its work cut out. Polls suggest that the Five Star Movement, founded by comedian Beppe Grillo and led by Luigi Di Maio, will receive the highest number of votes. Di Maio has actively wooed academics, bringing some on board as advisers. But most researchers regard the movement with alarm. Some of its members have vociferously supported anti-science campaigns, including that against vaccination.

The movement is unlikely to take part in any governing coalition. So the most likely government to emerge will be a mix of centre-right parties led by Silvio Berlusconi's Forza Italia and including the regionalist League, which is expected to receive the second highest number of votes. But whatever the content of the next government, says Mattia Butta, an Italian engineer at the Czech Technical University in Prague, it is unlikely to fundamentally change the scientific culture. ■

NUCLEAR PHYSICS

Physicists plan first antimatter road trip

Elusive material will be used to probe radioactive nuclei.

BY ELIZABETH GIBNEY

Antimatter is notoriously volatile, but physicists have learned to control it so well that they are now starting to harness it as a tool for the first time. In a project that began last month, researchers will transport antimatter by truck and then use it to study the strange behaviour of rare radioactive nuclei. The work aims to provide a better understanding of fundamental processes inside atomic nuclei, and to help astrophysicists learn about the interiors of neutron stars, which contain the densest form of matter in the Universe.

"Antimatter has long been studied for itself, but now it is mastered well enough that people can start to use it as a probe for matter," says Alexandre Obertelli, a physicist at the Technical University of Darmstadt in Germany, who leads the project, known as PUMA (antiProton Unstable Matter Annihilation), which will take place at CERN, Europe's particle-physics laboratory near Geneva, Switzerland.

CERN's antimatter factory makes antiprotons — the rare mirror image of protons — by slamming a proton beam into a metal target, and then dramatically slowing the emerging antiparticles so that they can be used in experiments. Obertelli and his colleagues plan to use magnetic and electric fields to trap a cloud of antiprotons in a vacuum. They will then load this trap into a van and drive it a few hundred metres to the site of a neighbouring experiment, known as ISOLDE, which produces rare, radioactive atomic nuclei that decay too quickly to be transported anywhere themselves. "It's almost science fiction to be driving around antimatter in a truck," says Charles Horowitz, a theoretical nuclear physicist at Indiana University Bloomington. "It's a wonderful idea."

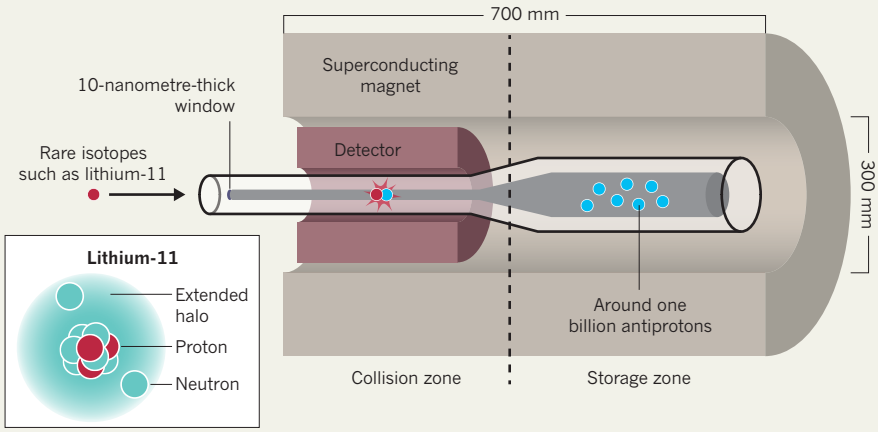
UNIQUE PROBE

Because antiprotons annihilate so readily, both with protons and with neutrons, they present a unique way to study the unusual configurations of radioactive nuclei. Whereas everyday atomic hearts host protons and neutrons in roughly equal measure, radioactive isotopes are stuffed with extra neutrons. This imbalance can give rise to exotic characteristics, including a surface 'skin' that is richer in neutrons than protons, or an extended halo in which neutrons orbit alone, as in lithium-11 (see 'Antimatter to go'). By observing how often antiprotons annihilate with a proton versus a neutron, the team will be able to understand the relative densities of these particles at the very edge of the nucleus. "It's a kind of test we haven't been able to do before on these new, more exotic nuclei, which may have very interesting structures," says Horowitz.

Radioactive nuclei act as microcosms for learning about neutron stars, objects that squash more mass than is contained in the Sun into the size of a city, and which are key to understanding how the Universe's heavy elements form. The cores of these

ANTIMATTER TO GO

To reveal the surface structure of atomic nuclei, physicists send ions of rare isotopes into a bottle 700 millimetres long — where they annihilate with antiprotons stored in the trap.



SOURCE: PUMA

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

CORRECTION

In saying that everyday atomic hearts have equal protons and neutrons, the News story 'Physicists plan first antimatter road trip' (*Nature* **554,** 412–413; 2018) didn't take account of the fact that some elements, such as hydrogen and lithium, have uneven numbers of protons in their most abundant form.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

super-dense stars remain a mystery, but their structure is dictated by the same little-understood interaction that creates bizarre phenomena in neutron-rich nuclei. "One of the reasons understanding neutron skins and halos is so important is to make the most of astrophysical observations," says Panagiota Papakonstantinou, a nuclear physicist at the Institute for Basic Science in Daejeon, South Korea.

Obertelli and his collaborators hope to create a trap that can store a record 1 billion antiprotons — more than 100 times as many as can be stored in any existing experiment. Another difficulty will be keeping the particles for weeks at a time, something that has so far been achieved for no more than a few dozen antiparticles at once. This will mean storing them at 4 degrees above absolute zero, and in a vacuum comparable with that of intergalactic space. "It's a challenging project," says Chloé Malbrunot, an antimatter physicist at CERN. "But I do think it is feasible."

Developing and testing technology for the portable trap will take around four years, with the first measurements scheduled for 2022. If the method works, physicists could transport antimatter much farther afield, allowing other scientists who aren't involved in the six experiments huddled at CERN's antiproton source to study and use the elusive matter.

"As soon as they can demonstrate one billion protons and keep them for several weeks, then many more experiments will join, and people with new ideas will come," says Malbrunot. "I think it will really open up the field." ■

OCEANOGRAPHY

Sensor array provides new look at global ocean current

Findings hint at the complexity of North Atlantic currents that drive the world's weather.

BY JEFF TOLLEFSON

The North Atlantic Ocean is a major driver of the global currents that regulate Earth's climate, mix the oceans and sequester carbon from the atmosphere — but researchers haven't been able to get a good look at its inner workings until now. The first results from an array of sensors strung across this region reveal that things are much more complicated than scientists previously believed.

Some research suggests that climate change is slowing currents in this area, which could affect weather and oceans around the world. Scientists with the Overturning in the Subpolar North Atlantic Program (OSNAP) are closer to pinning this down now. With data collected from 2014 to 2016, researchers have found that the strength of the Atlantic meridional overturning circulation — which pumps warm surface water north and returns colder water at depth — varies with the winds and the seasons, transporting an average of about 15.3 million cubic metres of water per second. The team presented its findings last week at an ocean-sciences meeting in Portland, Oregon.

The measurements are similar in magnitude to those from an array called RAPID, which has been operating between Florida and the Canary Islands since 2004. But scientists say they were surprised by how much the currents measured by the OSNAP array varied over the course of two years.

"We're delighted with the data," says Susan Lozier, an oceanographer at Duke University in Durham, North Carolina who is coordinating OSNAP. "We always find more complexity and more variability, but perhaps that shouldn't be surprising given how undersampled the ocean is."

MATHILDE JUTRAS



The research vessel *Polarstern* makes its way through sea ice off eastern Greenland.

OSNAP includes 53 moorings festooned with instruments that measure water conditions including temperature, salinity and current speeds. The network extends from southern Labrador in Canada to the southern tip of Greenland and on to the Rockall Trough off the west coast of Scotland (see 'Sea sensors'). And it's positioned where surface waters travelling north begin to cool and sink into the deep sea before turning south towards the Equator.

The array was designed to help answer fundamental questions about what drives this overturning circulation, which transports heat from the tropics to the poles. Climate models suggest that the formation of cold, deep water in the north might slow down as the climate warms. This would affect weather patterns in northern Europe and beyond, and some research suggests that this process could already be under way (S. Rahmstorf *et al. Nature Clim. Change* **5,** 475–480; 2015).

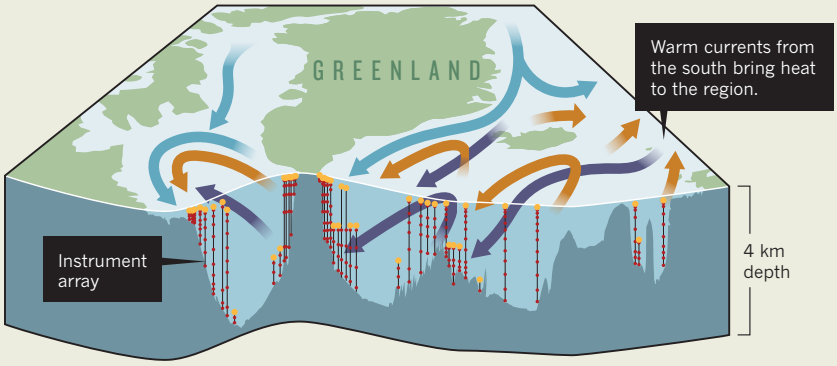
The RAPID array has registered a decline in the strength of this overturning circulation since 2008, including a 30% drop ▶

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

SEA SENSORS

Instruments placed across the North Atlantic are starting to reveal the inner workings of a global current that mixes the oceans.

Warmer surface currents ■ Cold and dense deep water ■ Cold water from the Arctic



SOURCE: OSNAP

from 2009 into 2010. A study published in *Geophysical Research Letters* on 12 February suggests that this slowdown is part of a longer-term pattern of climate variability across the north Atlantic, and not caused by global warming (D. A. Smeed *et al. Geophys. Res. Lett.* http://doi.org/cknk; 2018). Teasing a potential climate-change signal out of the data could take time, in part because the natural swings in the system are so large, says lead author David Smeed, an oceanographer at the National Oceanography Centre in Southampton, UK, who is involved with RAPID.

Researchers can't fully understand this slowdown without knowing what's happening in the North Atlantic, which drives the system. OSNAP is starting to fill in that knowledge gap.

Lozier hopes that the array will eventually settle a long-standing question about what powers this system of currents: how much of the overturning circulation is driven by the winds and how much by the formation of cold, dense water that sinks into the deep sea.

Since the array started collecting data, the North Atlantic has experienced two cold winters in a row. The data indicate that the ocean formed a lot of chilly water as a result. But it remains unclear to what extent this actually had an impact on the strength of the overturning circulation.

Eventually, Lozier wants to see how the overturning circulation responds during a warmer winter that produces less cold water in the region. "We just need to keep measuring," she says. ■

GENOMICS

CRISPR hack transforms cells into data recorders

Gene-editing tool can be harnessed to give a close-up view of life's most basic processes.

BY HEIDI LEDFORD

Researchers have harnessed the popular gene-editing tool CRISPR–Cas9 to turn DNA into a sensitive recording device. The recorder can document the duration and order of events within cells — and even erase and re-record information in the same genome.

The work, published in *Science* on 15 February[1], joins a flock of other CRISPR-based cellular recorders that have emerged from research labs over the past few years. The hope is that such recorders could keep track of changes in gene expression, trace a cell's individual family lineage or monitor shifts in environmental conditions.

"It's clear that the genome has this enormous recording capacity," says Jan Philipp Junker, a systems biologist at the Max Delbrück Center for Molecular Medicine in Berlin. "With CRISPR, we finally have the tools to use it."

It is one of many ways in which researchers are tinkering with the machinery of CRISPR–Cas9 to forge new types of molecular tool. Two papers also published in *Science* on 15 February describe how CRISPR can be used to detect disease-causing viruses[2,3].

The inspiration for the cellular recorders was the flight-data recorder found in many aeroplanes, says chemical biologist David Liu of the Broad Institute of MIT and Harvard in Cambridge, Massachusetts. "Just like a flight-data recorder records events that happen to an airplane, the cell-data recorders can be used to monitor stimuli that the cell is exposed to, or changes in the cell signalling," he says.

Researchers typically use CRISPR–Cas9 to alter DNA sequences by directing the Cas9 enzyme to cut DNA at a site dictated by the

MORE ONLINE

NATURE BRIEFING



Save time — get the daily *Nature* newsletter free in your inbox go.nature.com/2cxbubr

NEWS

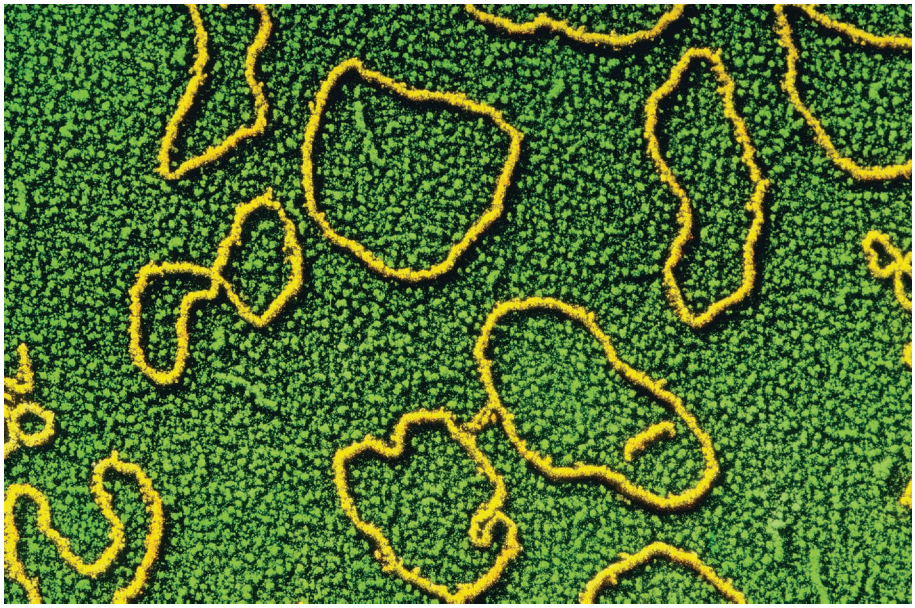
- Household chemicals such as soaps, paint and perfumes rival vehicles as a smog source go.nature.com/2hwr052
- Ocean tides could have driven ancient fish to walk go.nature.com/2gqfzwf

NATURE PODCAST

Defining adolescence; high-school researchers; and the science of teen risk taking nature.com/nature/podcast

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

sequence of a short fragment of RNA, called a guide RNA. In many organisms, these DNA breaks are then repaired by the cell in ways that can alter the original DNA sequence.



GOPAL MURTI/SPL

Circular DNA molecules called plasmids have been transformed into cellular-data recorders.

SLICE AND DICE

Liu and his colleague, chemist Weixin Tang, also at the Broad Institute, took advantage of Cas9's DNA-slicing ability to engineer a cellular recorder using circular molecules of DNA called plasmids. Plasmids replicate inside bacterial cells, sometimes producing hundreds of copies in a single cell.

Liu and Tang altered three DNA letters in one such plasmid, so that it contained a sequence targeted by a guide RNA. They also engineered a bacterium to express Cas9 only in the presence of a particular antibiotic, and dubbed the whole system CAMERA1.

Bacteria lack some of the DNA-repair mechanisms used by mammalian cells to fix the damage wrought by Cas9 — instead, when the plasmid is targeted by Cas9, it is degraded. Another plasmid then replicates to take the place of the lost one.

Liu and Tang put both the altered and normal plasmids into cells and measured the relative ratio of the two. The proportion of altered plasmid fell in cells that had been treated with antibiotic, because the cells had begun to degrade the altered plasmids.

The result was an unusually sensitive recorder: Liu and Tang could read out information from as few as ten bacterial cells. The magnitude of the change reflected the amount of antibiotic present, or the duration of exposure. Liu and Tang also developed methods for resetting the ratio of altered to unaltered plasmid, erasing the first recording and readying the cell to document the next event using the same set of plasmids.

They then moved on to create more recorders. One, called CAMERA2, relies on modified CRISPR systems called base editors, developed by Liu's lab[4] in 2016, that can alter single DNA letters in the genome without relying on the cell's DNA-repair mechanisms. Liu and Tang used CAMERA2 to record up to four different stimuli in bacterial cells, including exposure to light and viruses, and the order in which the stimuli occurred.

The researchers also modified CAMERA2 to work in mammalian cells, recording changes directly in the genome rather than in plasmids. Tang and her colleagues hope to use the system to answer questions about how cells take on specific identities.

The CAMERA recorders will join a host of other cellular recorders under development. Synthetic biologist Harris Wang of Columbia University Medical Center in New York City has been developing systems that he hopes to use to study microbes in the gut[5].

And Junker's lab is working on a recorder to trace the developmental pathway of individual cells in zebrafish. He first thought of the approach a little over two years ago, he says. "I thought this was an exotic idea that I had," says Junker. "And suddenly everybody is doing it." ■

1. Tang, W. & Liu, D. R. *Science* http://dx.doi.org/10.1126/science.aap8992 (2018).
2. Chen, J. S. *et al. Science* http://dx.doi.org/10.1126/science.aar6245 (2018).
3. Gootenberg, J. S. *et al. Science* http://dx.doi.org/10.1126/science.aaq0179 (2018).
4. Komor, A. C. *et al. Nature* **533,** 420-424 (2016).
5. Sheth, R. U., Yim, S. S., Wu, F. L. & Wang, H. H. *Science* http://dx.doi.org/10.1126/science.aao0958 (2017).

FUNDING

Indonesian funding agency short on cash

First dedicated fund has failed to raise enough money to finance projects.

BY DYNA ROCHMYANINGSIH

Indonesian scientists expecting to receive hundreds of thousands of dollars for research are stuck in limbo because the country's first dedicated science-funding agency is struggling to raise enough money.

After years of insufficient research funding, scientists were delighted when the Indonesian Science Fund (ISF) launched, in March 2016. Its aim was to boost investment in research and improve research quality. The following May, it opened up applications for grants worth up to 1.5 billion rupiah (US$110,000) a year for 3 years. Scientists submitted 467 proposals.

Almost 18 months later, the ISF, which is managed by the Indonesian Academy of Sciences in Jakarta, says it has selected 10 projects that deserve to be funded, but does not have the money to give to researchers.

"The delay from ISF is very disappointing," says a researcher in Jakarta who was told in February 2017 that their application might be accepted, but was asked to revise their project budget. The scientist, who asked to remain anonymous for fear of jeopardizing the payment, says that they have not heard from the ISF since. "I am trying to find international research funding. But as long as we don't get any official rejection from ISF, it is unethical for us to send [our] research proposal to other funding agencies," they say. "In this situation, scientists are the victims."

Although the ISF seeks funds from a range of sources, including philanthropy and ▶

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



DARREN WHITESIDE/REUTERS

Indonesian scientists were hopeful a dedicated fund would reverse years of insufficient investment.

foreign donors, the Indonesian government committed US$3 million to the agency in its first year, through the finance ministry's Indonesia Endowment Fund for Education. The endowment fund also signed a memorandum of understanding with the ISF, agreeing to give the agency money for five years.

Teguh Rahardjo, the ISF's executive director, says that the agency has received some of this cash, but not enough to support all of its grant programmes. Government budget constraints have contributed to the funding shortfall, and the agency is now searching for other sources of money, he says.

He says that the ISF was able to contribute 12 billion rupiah this year towards research grants for collaborations between Indonesian and UK universities. The programme is also supported by the Newton Fund, a British research-development group.

Indonesian researchers have typically had to rely on yearly funding from the national budget, which invests only 0.08% of the country's gross domestic product in scientific research each year. Some scientists argue that the process by which this money is awarded to universities or research centres by various ministries, including the Ministry of Research, Technology and Higher Education, isn't open and transparent. Syarif Hidayat, head of legal, cooperation and information services at the ministry, denies that the funding process is not open and says the ministry is working on ways to stretch research funding over multiple years.

The ISF was established to support long-term research. It is inspired by funding agencies such as the US National Science Foundation, and uses independent peer review to award funding.

"It is very unfortunate that ISF postpones this grant," says molecular biologist Riza Putranto at the Indonesian Research Institute for Biotechnology and Bioindustry in Bogor. "We cannot rely on national budget alone to conduct high-quality research in Indonesia," says Putranto, whose ISF funding application was not successful.

The Ministry of Finance did not respond to *Nature*'s questions about budget constraints. ■

CORRECTION

The Editorial 'Restore justice in Turkey' (*Nature* **554,** 145; 2018) stated that the text-messaging system ByLock was available only through personal introduction. In fact, it was widely available and could have been downloaded by anyone.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

COMMENT

SOCIETY Is Steven Pinker too optimistic about our troubled times? **p.420**

BIOLOGY From fly room to fly boom: how labs filled up with *Drosophila* **p.421**

EARTH OBSERVATION History shapes where, when, why and how we collect data **p.423**

METRICS Abolish author position — it is too crude a guide to contribution **p.423**



ZOONAR GMBH/ALAMY

Efforts to replicate research studies are distorted by inherent conflicts between the authors of the original work and those trying to reproduce the results.

How to make replication the norm

The publishing system builds in resistance to replication. **Paul Gertler, Sebastian Galiani** and **Mauricio Romero** surveyed economics journals to find out how to fix it.

Replication is essential for building confidence in research studies[1], yet it is still the exception rather than the rule[2,3]. That is not necessarily because funding is unavailable — it is because the current system makes original authors and replicators antagonists. Focusing on the fields of economics, political science, sociology and psychology, in which ready access to raw data and software code are crucial to replication efforts, we survey deficiencies in the current system.

We propose reforms that can both encourage and reinforce better behaviour — a system in which authors feel that replication of software code is both probable and fair, and in which less time and effort is required for replication.

Current incentives for replication attempts reward those efforts that overturn the original results. In fact, in the 11 top-tier economics journals we surveyed, we could find only 11 replication studies — in this case, defined as reanalyses using the same data sets — published since 2011. All claimed to refute the original results. We also surveyed 88 editors and co-editors from these 11 journals. All editors who replied (35 in total, including at least one from each journal) said they would, in principle, publish a replication study that overturned the results of an original study. Only nine of the respondents said that they would consider publishing a replication ▶

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

study that confirmed the original results.

We also personally experienced antagonism between replicators and authors in a programme sponsored by the International Initiative for Impact Evaluation (3ie), a non-governmental organization that actively funds software-code replication. We participated as authors of original studies (P.G. and S.G.) and as the chair of 3ie's board of directors (P.G.).

In our experience, the programme worked liked this: 3ie selected influential papers to be replicated and then held an open competition, awarding approximately US$25,000 for the replication of each study[4]. The organization also offered the original authors the opportunity to review and comment on the replications. Of 27 studies commissioned, 20 were completed, and 7 (35%) reported that they were unable to fully replicate the results in the original article. The only replication published in a peer-reviewed journal[5] claimed to refute the results of the original paper.

Despite 3ie's best efforts, adversarial relationships developed between original and replication researchers. Original authors of five of the seven non-replicated studies wrote in public comments that the replications actively sought to refute their results and were nitpicking.

One group stated that the incentives of replicators to publish "could lead to overstatement of the magnitude of criticism" (see go.nature.com/2gecz3b). Others made similar points. Although one effort replicated all the results in the original paper, the originating authors wrote, "we disagree with the unnecessarily aggressive tone of some statements in the replication report (particularly in the abstract)" (see go.nature.com/2esdjkr). Another group felt that "the statement that our original conclusions were robust was buried" (see go.nature.com/2siufet).

The 3ie-sponsored replication[5] of one highly cited paper[6] resulted in an acrimonious debate that became known as the Worm Wars. Several independent scholars speculated that assumptions made by the replicators had more to do with overturning results than with any scientific justification.

REPLICATION COSTS

A first step to getting more replications is making them easier. This could be done by requiring authors to publicly post the data and code used to produce the results in their studies — although this in itself would not redress the incentive to gain a publication by overturning results.

Ready access to data and code would ease replication attempts. In our survey of journal websites, the mid-tier economics journals (as ranked by impact factor) and those in sociology and psychology rarely asked for these resources to be made available. By contrast, almost all of the top-tier journals in economics have policies that require software code and data to be made available to editors before publication.

This is also true of most of the political-science journals we assessed, and all three of the broad science journals in our survey (see Supplementary information). In addition, many of the top-tier economics journals explicitly ask authors to post raw data as well as estimation data — the final data set used to produce the results after data clean-up and manipulation of variables. These are usually placed on a publicly accessible website that is designated or maintained by the journal. There seems to be no such norm in psychology and sociology journals (see 'Data checked?').

To see how often the posted data and code could readily replicate original results, we attempted to recreate the tables and figures of a number of papers using the code and data provided by authors. Of 415 articles published in 9 leading economics journals in May 2016, 203 were empirical papers that did not contain proprietary or otherwise restricted data. We checked these to see which sorts of files were downloadable and spent up to four hours per paper trying to execute the code to replicate the results (not including code runtime).

We were able to replicate only a small minority of these papers. Overall, of the 203 studies, 76% published at least one of the 4 files required for replication: the raw data used in the study (32%); the final estimation data set produced after data cleaning and variable manipulation (60%); the data-manipulation code used to convert the raw data to the estimation data (42%, but only 16% had both raw data and usable code that ran); and the estimation code used to produce the final tables and figures (72%).

The estimation code was the file most frequently provided. But it ran in only 40% of these cases. We were able to produce final tables and figures from estimation data in only 37% of the studies analysed. And in only 14% of 203 studies could we do the same starting from the raw data (see 'Replication rarely possible'). We do not think that spending significantly more time would have boosted our success — usually the code called for data or variables that had not been supplied.

Tellingly, the tables and figures could almost always be replicated if the code ran without major modifications. This was a

DATA CHECKED?

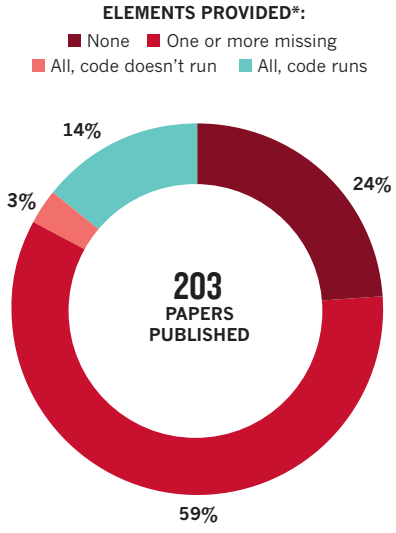
In a survey of 67 journals, most of the political-science and top-tier economics titles required authors to submit software code and data to editors before publication. Journals in sociology and psychology rarely did so.



SOURCE: P. GERTLER, S. GALIANI & M. ROMERO (UNPUBLISHED DATA)

REPLICATION RARELY POSSIBLE

An analysis of 203 economics papers found that fewer than one in seven supplied the materials needed for replication.



*The elements assessed were raw data, raw code, estimation data and estimation code.

SOURCE: P. GERTLER, S. GALIANI & M. ROMERO (UNPUBLISHED DATA)

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

significant hurdle: only 54% and 61% of articles had data-manipulation code or estimation code, respectively, that did not require major modifications.

Our results align with previous findings in the literature[2]. A study of the articles published in the *Journal of Money, Credit and Banking* found that only 35.7% of articles met data-archive requirements, and only 20% of studies could be replicated using the information available in the archive[7]. Another study attempted to replicate 67 papers published in 13 well-regarded general-interest and macroeconomics journals, and could replicate only 29 (ref. 8). This problem goes beyond economics. In 2013, only 18 of 120 political-science journals had replication policies[9], and a 2016 study found that only 58% of the articles in top political-science journals publish their data and code[10].

However, some progress is being made. Notably, a few political-science journals verify that posted data and code produce the results in a publication. A handful of journals in statistics and information sciences have appointed 'reproducibility editors' to ensure that analyses can be replicated. The American Economic Association appointed a data editor last year to oversee reproducibility in its journals.

"Usually the code called for data or variables that had not been supplied."

CREDIBLE THREAT

We think that the way forward is for more journals to take on this kind of responsibility, using data editors to help implement the following replication policy.

Journals could oversee the replication exercise after conditional acceptance of a manuscript but before publication. Journals would then verify that all raw data used in the paper and code (that is, sample and variable construction, as well as estimation code) are included and executable. They would then commission academic experts, advanced graduate students or their own staff to verify that the code reproduces the tables and figures in the article. If not, editors could ask authors to correct their errors and, if necessary, have papers reviewed again.

In addition, for a random sample of papers, journals should attempt to reconstruct the code from scratch or search the executable code for errors. In this way, all papers would have some positive probability of being fully replicable.

This simple procedure has the winning combination of four desirable characteristics. First, it is unbiased: editors would have no incentive to overturn results. Second, it creates a credible expectation in authors that their work will be replicated, motivating them to be careful and to put effort into constructing their code and not report false results. Third, data and code would be easily available to outside researchers to explore the robustness of the original results using alternative specifications, measurements and methods of identification and estimation. Fourth, there is little cost associated with getting a research associate to perform replication exercises, especially because authors have strong incentives to cooperate with journals at the pre-publication stage. Although we are writing from our experience as economists, we think that similar practices could be adopted in other disciplines, particularly the social sciences.

Initially, these steps might slow down the time from acceptance to publication for some papers. However, authors will eventually internalize them and submit accurate, error-free materials so that the study replication will be done efficiently. This will help to restore confidence in the credibility of science. ■

Paul Gertler *is a professor of economics in the Haas School of Business at the University of California, Berkeley, USA.* **Sebastian Galiani** *is a professor of economics at the University of Maryland, College Park, USA.* **Mauricio Romero** *is a graduate student in economics at the University of California, San Diego, USA.*
e-mail: gertler@berkeley.edu

1. Nosek, B. A. *et al*. *Science* **348,** 1422–1425 (2015).
2. Christensen, G. & Miguel, E. Preprint at http://dx.doi.org/10.17605/OSF.IO/9A3RW (2017).
3. Berry, J., Coffman, L. C., Hanley, D., Gihleb, R. & Wilson, A. J. *Am. Econ. Rev.* **107,** 27–31 (2017).
4. International Initiative for Impact Evaluation. *3ie Replication Programme* (3ie, 2017); available at http://go.nature.com/2eoe7o5
5. Davey, C., Aiken, A. M., Hayes, R. J. & Hargreaves, J. R. *Int. J. Epidemiol.* **44,** 1581–1592 (2015).
6. Miguel, E. & Kremer, M. *Econometrica* **72,** 159–217 (2004).
7. McCullough, B. D, McGeary, K. A. & Harrison, T. D. *J. Money Credit Bank.* **38,** 1093–1107 (2006).
8. Chang, A. C. & Li, P. *Finance Econ. Disc. Ser.* http://dx.doi.org/10.17016/FEDS.2015.083 (2015).
9. Gherghina, S. & Katsanidou, A. *Eur. Polit.Sci.* **12,** 333–349 (2013).
10. Key, E. M. *Polit. Sci. Politics* **49,** 268–272 (2016).

Supplementary information and funding details accompany this article online; see go.nature.com/2gejxbd.

CORRECTION

The Comment article 'Join the disruptors of health science' (*Nature* **551,** 23–26; 2017) should have disclosed that *Nature*'s Editor-in-chief, Philip Campbell, serves on an unpaid basis as a member of the science advisory board for the start-up firm Mindstrong Health.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



BRIDGEMAN IMAGES

The Marseillaise **(1870) by Gustave Doré: the French Revolution began with ideals of rationalism, but descended into chaos and conflict.**

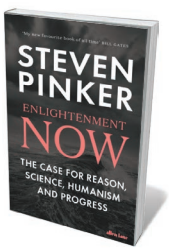
SOCIETY

Testing times for optimism

Ian Goldin questions Steven Pinker's faith in Enlightenment values in our complex era.

Steven Pinker's *Enlightenment Now* builds on his 2011 *The Better Angels of Our Nature* (Viking) in offering another engaging, compelling set of reasons to be cheerful. In this new combined survey, analysis and manifesto, he convincingly demonstrates that when it comes to health and life expectancy, poverty reduction and income, education, human rights, peace and security, the global data provide solid grounds for optimism. But the book's premise lies in the past: the Enlightenment, that period in the eighteenth century when, Pinker argues, reason, science, humanism and progress became the centre of intellectual endeavour in Europe and North America. That legacy, he asserts, is ripe for resurrection at a time of political upheaval, the rise of demagoguery, climate scepticism and 'fake news'.

The Enlightenment undoubtedly saw major advances in constitutional government. The separation of church and state in some nations allowed new models of society to flourish. But using the era as a premise is problematic. Many of the breakthroughs that Pinker attributes to the Enlightenment actually pre-date it. As Chris Kutarna and I showed in *Age of Discovery* (Bloomsbury, 2017), the Renaissance was a period of even more dramatic progress in science and the humanities, sparked by luminaries such as the astronomer Nicolaus Copernicus and the humanist Erasmus (see P. Ball *Nature* **452,** 816–818; 2008).



Enlightenment Now: The Case for Science, Reason, Humanism, and Progress
STEVEN PINKER
Viking: 2018.

Before that were extraordinarily innovative epochs in Asia and other regions, such as China's Tang dynasty (AD 618–907) and the Islamic Golden Age (750–1260).

Another issue we highlighted is that less salubrious aspects of change accompanied eras of rapid progress. During the Renaissance voyages of discovery, European sailors spread diseases that killed millions of indigenous Americans. Europe saw the rise of religious extremists — such as the demagogue Girolamo Savonarola, who ruled Florence in the 1490s — along with intolerance of diversity and science. Enlightenment advances were tied to empire building and the nascent Industrial Revolution, predicated not just on noble ideas and scientific curiosity, but also on slavery, genocide, exploitation and cultural triumphalism. Like the Renaissance, it ended in chaos and conflict, including the French revolutionary wars of 1792–99.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

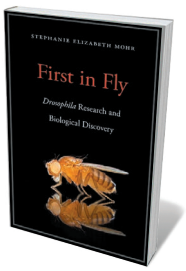
Both eras show that science and evidence-based thinking do not necessarily triumph over irrationality and ideology. Shared social norms and ethics are the framework that allows reason to prevail. Moreover, progress in science and society is a tapestry that cannot be unpicked through simple periodization. That said, comparing historical eras can be instructive. Although history does not repeat itself, it may rhyme; it can offer insights into predicaments and solutions, as I have found helpful, for instance, in conceptualizing the future of work. Many of the Enlightenment's philosophical, social and scientific advances offer lessons for our reason-starved times, as Pinker ably argues; but context is key.

Although it is framed as a historically informed template for a new age of reason, *Enlightenment Now* ultimately becomes something else: an extended dismissal of the arguments of despair that Pinker fears are defining politics and crowding out an alternative approach rooted in rationality and global cooperation. He does not frame the thesis in economic terms. Yet he essentially defends globalization and the growth of market economies by claiming that it has brought more progress than any force in history. As an economist, I agree.

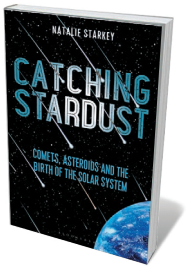
But globalization has also led to an escalation of risks. What is rational for individuals is increasingly irrational for society. The drivers of progress are rising incomes and connectivity; these also lead to greater negative spillovers and systemic risk. Managing globalization's underbelly is essential, and the gulf between what needs to be done and what is being done is widening. Economic growth has come at the expense of ecosystems. Because nature does not respond to price signals (rhinos do not reproduce more when their horns are more valuable), increasing freedom of choice has led to overexploitation of a growing number of natural systems. Pinker does cite climate change, but as a worrying exception to a relentlessly positive narrative, rather than as the most glaring example of a wider failure of global commons management.

Pinker devotes a chapter to inequality. Drawing on the encyclopaedic website Our World in Data by economist Max Roser of the Oxford Martin School, UK, he shows that global inequality (between richer and poorer countries) is decreasing, whereas inequality within countries is rising. He concludes that much current anxiety around inequality — spurred by the idea that the top 1% is absorbing the lion's share of economic growth — is misplaced. I don't agree. First, the data he presents do not reflect recent sharp economic reversals in many places. These are likely to become more significant in some regions. For instance, in the United States, cuts to healthcare subsidies and corporate tax rates over the past few months will reduce redistribution ▶

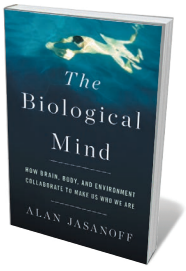
Books in brief



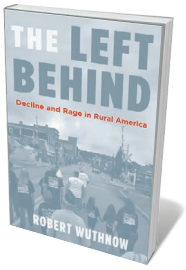
First in Fly: *Drosophila* Research and Biological Discovery
Stephanie Elizabeth Mohr HARVARD UNIVERSITY PRESS *(2018)*
How did an inconspicuous fly, *Drosophila melanogaster*, become the laboratory "model of models"? Since the early twentieth century, when pioneering geneticist Thomas Hunt Morgan created the Fly Room at Columbia University in New York City, the insect's brief life, vast broods, obvious mutant phenotypes and genetic similarities to humans have made it a boon for the field. In this admirable study, *Drosophila* researcher Stephanie Elizabeth Mohr reveals a raft of breakthroughs discovered "first in fly", such as the molecular mechanisms of circadian rhythm. Serious science, elegantly described.



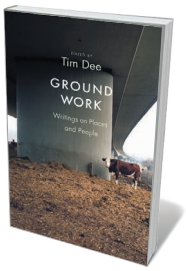
Catching Stardust
Natalie Starkey BLOOMSBURY SIGMA *(2018)*
As the runts of our rock-ridden cosmic neighbourhood, comets and asteroids might seem predictable. Not so, proves space geologist Natalie Starkey in this pop-science dash through findings on the denizens of the Kuiper belt and beyond. Starkey is perhaps most eloquent on comets, those icy scoops of early solar-nebula cloud. The European Space Agency's audacious Rosetta mission, for instance, revealed comet 67P to be dark, dusty, rubber-duck-shaped and redolent with an unusual bouquet of compounds. Space mining and strategies for tackling an asteroid hit get a look-in, too.



The Biological Mind
Alan Jasanoff BASIC *(2018)*
In this powerful treatise, neurological engineer Alan Jasanoff issues a corrective to the "cerebral mystique" — the idea that the human brain is a machine-like outlier in the body. As he reminds, the brain is deeply interconnected, and functions using chemicals, electricity, neurons, glia, passive diffusion and active signalling. This integrated view of brain, body and mind could, he asserts, revolutionize discourse on mental illness: the 'broken-brain' model could give way to one that meshes genetic, environmental and cultural factors, thereby adding needed nuance to our understanding of causes and treatment.

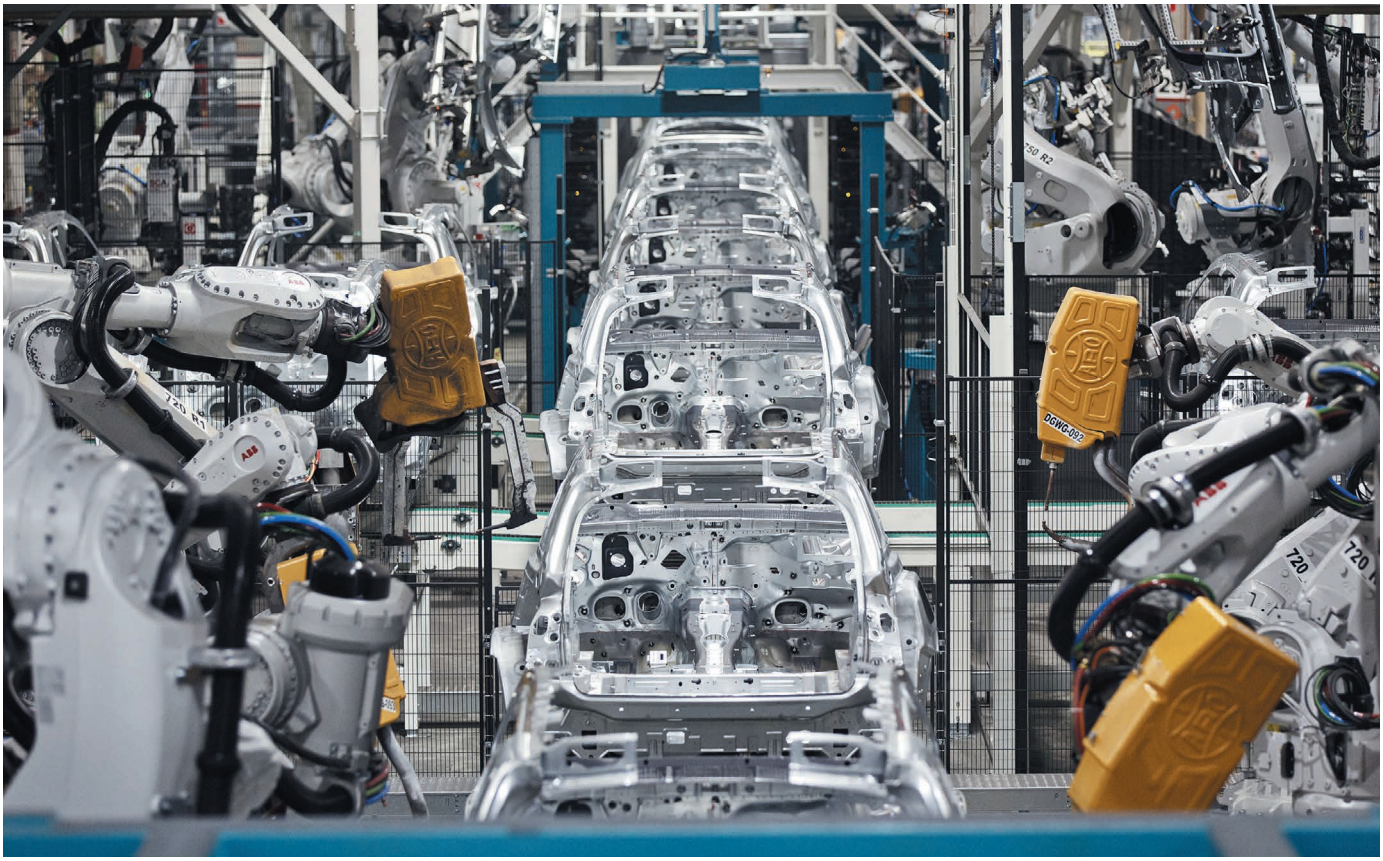


The Left Behind: Decline and Rage in Rural America
Robert Wuthnow PRINCETON UNIVERSITY PRESS *(2018)*
Ever since 62% of the rural vote went to Donald Trump in the 2016 US presidential race, rage in the heartlands has become a media trope. In this thoughtful study — distilling a decade of research and more than 1,000 interviews — social scientist Robert Wuthnow digs beyond the "grievance-and-resentment" stereotype of this 50-million-strong, overwhelmingly white population. He concludes that the rural population should be seen as "moral communities": united by mutual obligation and cultural norms, but threatened by the alien, whether local drug abuse or the culture of Washington DC.



Ground Work
Edited by Tim Dee JONATHAN CAPE *(2018)*
This superb anthology is a paean to spirit of place in dislocated times. With UK eco-charity Common Ground, Tim Dee has gathered personal geographies from 31 poets and writers. And it is a trove. Sean Borodale likens Somerset honeycomb to a "microfiche" storing local data on flora and weather; Julia Blackburn sees Suffolk shells on a table looking "like music, or a story without the need of words"; and Hugh Brody recalls how, through mapping, he discovered the intimate relationship of the Inuit with the Arctic. **Barbara Kiser**

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



BLOOMBERG/GETTY

Robots construct cars at a Mercedes-Benz plant in Finland.

and weaken the safety nets that (as Pinker identifies) helped to alleviate inequality after the 2008 financial crisis.

Second, he does not give enough weight to the plight of the jobless — and, in the United States, its association with the growing problem of opioid misuse — or the threat that robots and artificial intelligence (AI) pose to jobs. Growing inequality within countries is undermining social cohesion and promoting extremist political views, and this is likely to be compounded by technological change.

If most tasks that are repetitive and do not demand dexterity are automated — which is possible, according to research at the Oxford Martin School — what is the model of future employment in administrative back offices, call centres or manufacturing? Pinker is too sanguine about such immediate threats (although he argues cogently in favour of minimizing the expenditure of intellectual and political capital on remote risks, such as the existential threat posed by AI). He also underestimates the threats that Mike Mariathasan and I set out in our 2015 book *The Butterfly Defect* (Princeton University Press): interconnected systems raise interdependence, complexity and systemic risk. With globalization, as connectivity, incomes and consumption grow, so risks become multidimensional and transnational — witness pandemics, cascading financial crises and cyberattacks. New technologies turbocharge these risks and give new powers to individuals and small groups.

Pinker looks in some depth at the rise in scepticism about science, in society and in parts of academia such as the humanities. He gives shorter shrift to the erosion of faith in expertise elsewhere, even though this is justified in some notable instances. The financial sector, for example, is home to the biggest and most powerful global expert system, from banks and treasuries to the International Monetary Fund. The 2008 financial crisis highlighted the inadequacy of that system, whose primary mission is financial stability. Similarly, accounting firms have given clean bills of health to corrupt or collapsing enterprises. The denial of evidence is irrational, but it is not necessarily irrational to challenge experts and authorities.

> ***"The denial of evidence is irrational; but it is not necessarily irrational to challenge experts."***

Pinker highlights the importance of ideas in human advances, but claims that "intellectuals hate progress" — a surprising generalization in a case for evidence and reason. And herein lies both the strength and weakness of this book. It provides copious reasons for optimism. It claims to be a scientific, evidence-based account; but that is undermined by its scant attention to the mounting evidence on new risks. I share Pinker's optimism that this could be our best century, in which poverty and many of the challenges humanity has historically confronted are addressed. Yet there is also a real potential for dystopian outcomes as sea levels, strife, temperatures and resistant infections rise, and biodiversity, democratic institutions, social ties, mental health and resource security are eroded. We need to face up to these and other daunting challenges while nurturing the positivity required to tackle them.

Enlightenment Now is not a balanced account of the present or future. But for the many overwhelmed by gloom, it is a welcome antidote. ■

Ian Goldin *is Professor of Globalisation and Development and director of the Oxford Martin Programme on Technological and Economic Change at the University of Oxford, UK, and the author of* Age of Discovery *and* The Butterfly Defect. *His website is https://iangoldin.org/. Twitter: @ian_goldin*

CORRECTION

The Books & Arts article 'Epiphanies of the edgelands' (*Nature* **554,** 166–167; 2018) incorrectly referred to East Greenland instead of West Greenland.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Correspondence

Society and history imprint climate data

Long-term climate observations that are highly processed can obscure the context in which the data were collected. Yet such background can provide important information about how climate and society interact, and could contribute to the design of climate services.

Many historical measurements were part of colonialism: describing and explaining the world according to European ideas. Polar explorations left their traces on climate records, and wars stimulated measurements in strategic locations while interrupting others or destroying archives. Cold-war geopolitics shaped modern global networks for collecting climate information and prompted the creation of World Data Centers to disseminate it. Trade and whaling were entwined with marine climate data. Restrictive policies can still impose omissions on data maps. And commercial aviation generates weather records from flight corridors.

In our view, such context dependence could inform climate services. In developing countries, for instance, farming practices that use traditional knowledge of local climates might need adjusting in response to global warming.

Scientists working with climate data currently try to minimize the effect of context on their estimates of physical variables. In our view, policy decision-making should be based on both.

Stefan Brönnimann, Jeannine Wintzer *University of Bern, Switzerland.*
stefan.broennimann@giub.unibe.ch

Roots out of sight, not out of mind

Using genomics and imaging to understand plants' physical traits does not stop above ground (see *Nature* **553,** 396–398; 2018). The architecture of plant root systems growing in opaque soil is being revealed by 3D-imaging tools such as X-ray computed tomography, magnetic resonance imaging and neutron tomography.

Combining these insights (see, for example, T. Roose *et al. Plant Soil* **407,** 9–38; 2016) with gene-expression studies, spatially resolved data on chemical and physical parameters, and the composition of the soil microbiome around roots should help to explain how system properties emerge as these components interact. Such properties are relevant to the acquisition of water and nutrients, as well as to plant health, carbon storage in soil and soil structure.

Priorities are to develop joint sampling protocols and a common language across physics, soil science and biology, along with skills in image analysis, spatial statistics and modelling at scales from root hairs to whole-root systems.

Germany's main research funder (the DFG) has just launched a programme to study the spatio-temporal organization and function of the soil zone altered by the plant root, also called the rhizosphere (see www.ufz.de/spp-rhizosphere).

Doris Vetterlein, Steffen Schlüter, Hans-Jörg Vogel *Helmholtz Centre for Environmental Research — UFZ, Halle, Germany.*
doris.vetterlein@ufz.de

Temperature debt of solar geoengineering

Solar geoengineering is a proposed method of climate engineering that aims to reduce global warming using an artificial 'sunscreen' of aerosols in Earth's high atmosphere. As planning of the first field experiments gets under way, any potential risks associated with the technology must be transparently assessed and not downplayed or dismissed. One such risk of solar geoengineering is its 'temperature debt' — the planetary heating that would arise if maintenance of the artificial sunscreen was discontinued.

Modelling suggests that most of the world's population could benefit from this temporary sunscreen, compared with the adverse effects of unabated climate change (a questionable reference state, in my view). However, models also reveal the alarming rise in temperatures that could occur if the screen of short-lived aerosols should suddenly cease to function for any reason in the presence of high concentrations of long-lived greenhouse gases (see H. D. Matthews and K. Caldeira *Proc. Natl Acad. Sci. USA* **104,** 9949–9954; 2007). This rapid warming would pose a severe risk to ecosystems and society.

Even with the best planning to ensure steady operation of the technology, its continuous safe functioning and maintenance cannot be guaranteed. Yet it could take hundreds of years to safely phase out solar geoengineering and achieve the same degree of cooling by reducing greenhouse-gas concentrations. It is therefore imperative that, in the absence of a fail-safe mode for solar geoengineering, the temperature debt is fully accounted for in any assessments of this technology.

Andreas Oschlies *GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany.*
aoschlies@geomar.de

Drop author position as a proxy metric

Marcus Munafò and George Davey Smith suggest that using different lines of evidence (triangulation) to verify results will change how credit is assigned to authors of research papers (*Nature* **553,** 399–401; 2018). In my view, it would help if we were to abolish the convention of using an author's position in the list as a proxy weighting of contribution. Such a move would shift the assessor's curiosity to the 'author contributions' section, and render evaluation more meaningful and transparent.

Currently, author sequence still claims more attention than short descriptions of author contributions in those journals that include them (D. van Dijk *et al. Curr. Biol.* **24,** R516–R517; 2014). If individual careers in many fields continue to depend on first and last authorship positions, scientific progress and even quality could become a lesser concern. And simple quantitative metrics, such as linear ranking of authors according to contribution, are open to abuse (P. E. Smaldino and R. McElreath *R. Soc. Open Sci.* **3,** 160384; 2016).

If it is only the specified contribution that counts, there will be a strong incentive for researchers to contribute effectively and in innovative ways. It would encourage collaboration and discourage honorary and 'guest' authorship. Indeterminate contributions would become more apparent.

As for the roll-call of authors, let's explore such alternative formats as alphabetical listing, which is often used in mathematics, economics and high-energy physics.

Daniel Hornburg *Stanford University, California, USA.*
hornburg@stanford.edu

CORRECTION

The Nature Index article 'Industry links boost research output' (*Nature* **552,** S11–S13; 2017) erroneously located Keyvan Vakili at the University of Southern California. He is actually at the London Business School.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

COMING OF AGE

A special issue explores the maturing science of adolescence.

IT'S WIDELY ACCEPTED THAT ADOLESCENTS ARE MISUNDERSTOOD. Less well known is how far we still have to go to understand adolescence itself. One problem is that it is hard to characterize: the concept of puberty does not capture the decade or more of transformative physical, neural, cognitive and socio-emotional growth that a young person goes through. Another is that science, medicine and policy have often focused on childhood and adulthood as the most important phases of human development, glossing over the years in between.

Yet understanding adolescents and adolescence is crucial. It's becoming clear that physical and mental health at adolescence often set a trajectory for the rest of life. Meanwhile, today's teenagers are projected to make up the largest generation in human history. Most of the world's 10- to 19-year-olds are maturing in parts of the globe that are facing enormous health, environmental and socio-economic challenges. And all are growing up in an ever-transforming digital world.

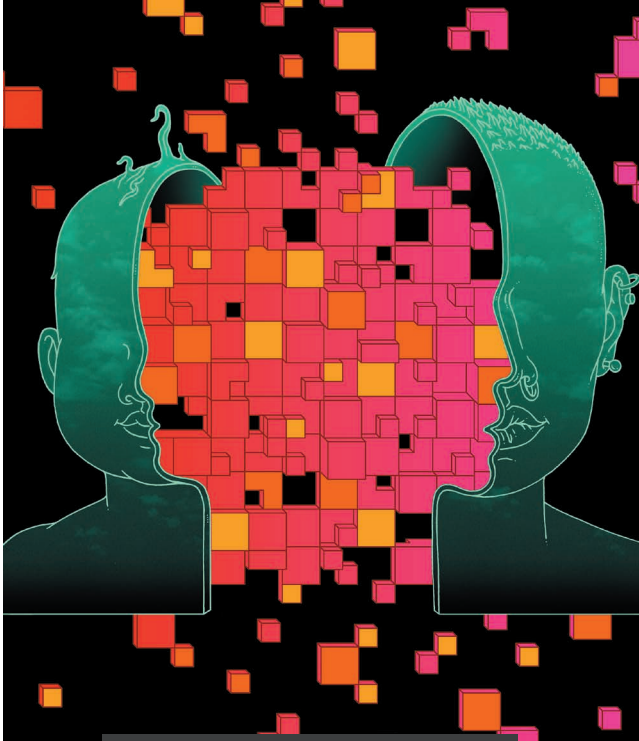
In this issue, we explore the science of adolescence — a field that is itself coming of age. A News Feature on page 429 explores how the boundaries of adolescence are widening, thanks to advancing puberty and shifting societal definitions. Another News Feature, on page 426, describes how neuroscientists are showing that risk-taking and rebellion are more nuanced than stereotypes suggest — and are sometimes even useful. There will be an accompanying podcast at go.nature.com/2geosm3 and a video at go.nature.com/2syvbdh.

In a Comment on page 432, developmental psychologist Candice Odgers argues that digital technology is not the widespread menace it is often made out to be — but that young people struggling in their offline lives may be at increased risk online. On page 435, social scientists Robert Blum and Jo Boyden stress the importance of understanding the daily lives of adolescents in low- and middle-income countries, home to 90% of the world's 10- to 24-year-olds. A News & Views on page 438 appraises the Feinberg hypothesis, which posits that abnormal pruning of neural connections in adolescence leads to schizophrenia, and a Careers Feature on page 559 looks at scientists who give teenagers jobs in the lab.

Three reviews explore some paradoxes of adolescence. On page 441, paediatrician Ronald Dahl and his colleagues review the evidence for adolescence as a physical, cognitive and socio-emotional growth spurt, and look at how to design policies that take into account teenagers' unique needs and values. On page 458, epidemiologist George Patton and his colleagues explore how factors that harm adolescent health can affect the next generation. And on page 451, anthropologists Carol Worthman and Kathy Trang use life-history theory to interpret shifts in biological puberty, and in cultural definitions of and influences on adolescence.

Nature's articles are part of a collection of content (http://nature.com/collections/adolescence) in nine journals from the Nature family and its sister publication, *Scientific American*. It is the largest such collection that the Nature journals have ever published — and one that we hope proves valuable for scientists, policymakers and parents alike. ■

ILLUSTRATION BY ERIC NYQUIST



CONTENTS

FEATURE

426 **Sex, drugs and self-control**
Kerri Smith

429 **The shifting boundaries of adolescence**
Heidi Ledford

COMMENT

432 **Smartphones are bad for some teens, not all**
Candice Odgers

435 **Understand young people in low-income countries**
Robert Blum & Jo Boyden

NEWS & VIEWS

438 **Pruning hypothesis comes of age**
Matthew B. Johnson & Beth Stevens

PERSPECTIVES

441 **Importance of investing in adolescence from a developmental science perspective**
Ronald E. Dahl, Nicholas B. Allen, Linda Wilbrecht & Ahna Ballonoff Suleiman

451 **Dynamics of body time, social time and life history at adolescence**
Carol M. Worthman & Kathy Trang

ANALYSIS

458 **Adolescence and the next generation**
George C. Patton et al.

CAREERS

559 **Teen spirit in the lab**
Chris Woolston



ADOLESCENCE
A *Nature* special issue
nature.com/collections/adolescence

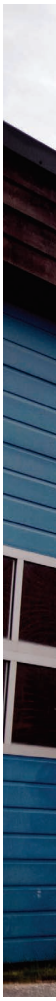
© 2018 Macmillan Publishers Limited, part of Springer Nature

SEX, DRUGS AND SELF-CONTROL

It's not just about rebellion.

Neuroscience is revealing adolescents' rich and nuanced relationship with risky behaviour.

BY KERRI SMITH



Cole Skinner was hanging from a wall above an abandoned quarry when he heard a car pull up. He and his friends bolted, racing along a narrow path on the quarry's edge and hopping over a barbed-wire fence to exit the grounds.

The chase is part of the fun for Skinner and his friend Alex McCallum-Toppin, both 15 and pupils at a school in Faringdon, UK. The two say that they seek out places such as construction sites and disused buildings — not to get into trouble, but to explore. There are also bragging rights to be earned. "It's just something you can say: 'Yeah, I've been in an abandoned quarry'," says McCallum-Toppin. "You can talk about it with your friends."

Science has often looked at risk-taking among adolescents as a monolithic problem for parents and the public to manage or endure. When Eva Telzer, a neuroscientist at the University of North Carolina in Chapel Hill, asks family, friends, undergraduates or researchers in related fields about their perception of teenagers, "there's almost never anything positive", she says. "It's a pervasive stereotype."

But how Alex and Cole dabble with risk — considering its social value alongside other pros and cons — is in keeping with a more complex picture emerging from neuroscience. Adolescent behaviour goes beyond impetuous rebellion or uncontrollable hormones, says Adriana Galván, a neuroscientist at the University of California, Los Angeles. "How we define risk-taking is going through a shift."

Adolescents do take more risks than adults, and the consequences can include injury, death, run-ins with the law and even long-term health problems. But lab studies in the past decade have revealed layers of nuance in how young people assess risks. In some situations, teenagers can be more risk-averse than their older peers. And they navigate a broader range of risks than has typically been considered in the lab, including social risks and positive risks — such as trying out for a sports team. These types of behaviour seem to have different effects on the brain.

How adolescents interact with risk is important. Work on the neural underpinnings of risky behaviour can inform guidelines and laws for teens who drive, for example, or the punishments they receive for violent crimes. Understanding how the teenage brain evaluates risk could even reveal predictors of mental-health conditions such as schizophrenia and depression, which often emerge in adolescence.

In more ways than one, there is a lot going on in a teenager's head. "In fact, it's just beautiful," says B. J. Casey, a neuroscientist at Yale University in New Haven, Connecticut. "It's amazing that it unfolds correctly most of the time."

REBEL WITH A CAUSE

Adolescence is a perilous period. The death rate among 15- to 19-year-olds worldwide is about 35% higher than that among 10- to 14-year-olds. And risky behaviours are linked to many of the major threats to life during this time (see 'Risking life and limb'). Road injuries are the biggest cause of death for adolescents globally. Self-harm and other forms of violence also rank highly. Plus, some practices that can lead to poor health in adulthood — such as use of tobacco or alcohol, or sedentary lifestyles — often stem from poor choices made in the teenage years. So, risky behaviour has been a preoccupation for scientists.

> "RISK-TAKING HAS DRIVEN A LOT OF THE EARLY WORK. IT WAS A ROUTE TO SUCCESSFUL FUNDING."

"Risk-taking has driven a lot of the early work" on the teenage brain, says Ronald Dahl, who studies adolescent brain development at the University of California, Berkeley. "It was a route to successful funding, so it was emphasized."

Early theories focused on a perceived imbalance in the developing brain. Areas linked with impulsivity and heightened sensitivity to reward, especially in the social realm, get an early boost in activity, whereas those governing cognitive processes such as working memory develop smoothly throughout adolescence.

Neuroscientists likened the emerging picture of the teenage brain to that of a car with a revving accelerator and faulty brakes. This fit the developmental data, but not the fact that many teenagers show no proclivity for risk-taking, says Ted Satterthwaite, a psychiatrist and neuroimaging researcher at the University of Pennsylvania in Philadelphia. A 2016

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Teenagers in Norway practise the urban sport parkour.

ANDREA GJESTVANG/PANOS

survey[1] of more than 45,000 US teenagers found that 61% had not tried cigarettes by age 17–18, for example; some 29% had never drunk alcohol.

Most neuroscientists now acknowledge that neural systems developing at different rates do not mean that the brain is unbalanced. "It's a vulnerable period, but it's not vulnerable just because there's something going wrong with their brains," says Satterthwaite.

And so work has shifted to looking at a broader range of risks and environmental influences. For many teenagers, says Dahl, there is risk in relatively benign experiences, such as standing up for a friend or asking someone on a date. "Taking a social risk — those feel more salient."

THE SOCIAL WHIRL

In recent years, studies have begun to characterize how social elements influence risk. In 2009, Laurence Steinberg, a psychologist at Temple University in Philadelphia, got teenagers to lie in a functional magnetic resonance imaging (fMRI) scanner and play 'the chicken game' — a video game in which they drive a car, passing through an implausible 20 traffic lights in 6 minutes. As the first lights change to amber, some teenagers choose to carry on; others wait for green. Sometimes speeding ahead pays off, but sometimes the car gets hit.

When teenagers played this game alone, they took risks at about the same frequency as adult players[2]. But when Steinberg told the adolescents that their friends were watching from an adjacent room, they took significantly more risks. In a similar study[3] by Telzer and her colleagues, teenagers took fewer risks when they were told that their mothers were watching. The scanner revealed greater activation in reward-sensitive brain regions, such as the ventral striatum, with the friend-influenced risky behaviours. Meanwhile, the mothers' presence correlated with activation in the prefrontal cortex, an area known to be involved in cognitive control.

Neuroscientists have used this game to test how a teenager's propensity to take risks can depend on their social stature. In one study[4], a team at the University of Oregon in Eugene got adolescents to play it in a scanner after hearing that two other teenagers were watching. Then the researchers got the participants to play another video game, in which they were excluded from throwing and catching a ball with the same peers.

When they returned to the driving game after experiencing social exclusion, adolescents who said they were sensitive to peer influence took significantly more risks. Those that demonstrated this pattern also showed greater activation in a brain area involved in modelling the thoughts of others, the temporoparietal junction. In another study[5], Telzer and her colleagues found that teenagers who were more socially excluded or victimized took more risks. The work is part of a drive to understand who is most vulnerable. "If we know the context under which teens smoke or make good or bad decisions, we can push them into the contexts that are more positive," says Telzer.

Peers can have positive effects, too. In a 2014 study[6], teenagers were asked to donate or keep money in an online game, supposedly watched by ten peers. If a participant made a donation and their peers approved — denoted by a 'thumbs up' icon — the participant made more donations during the game. (Although the opposite is also true.) "There's an assumption that teenagers' friends are a monolithic negative influence," says Telzer. The real picture is more complex.

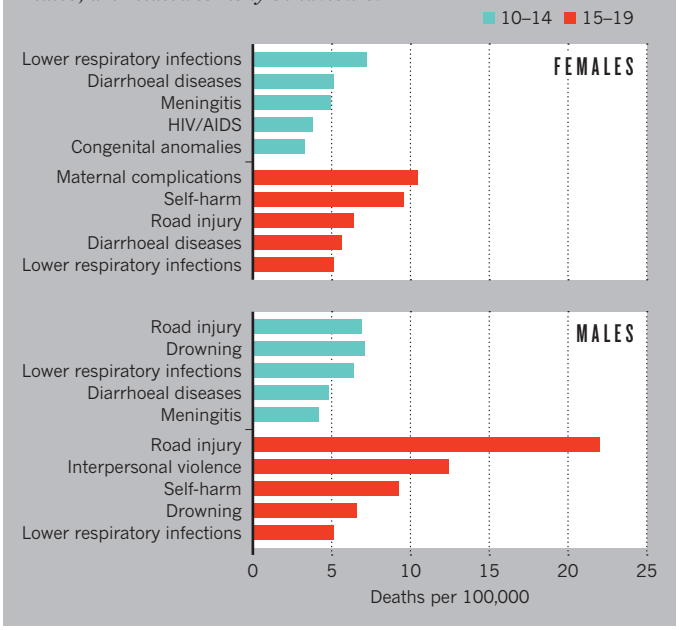
Interestingly, the same brain systems that mediate unhealthy risk-taking also seem to help teenagers to take positive risks. Activity in the ventral striatum, particularly rising numbers

ADOLESCENCE
A *Nature* special issue
nature.com/collections/adolescence

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

RISKING LIFE AND LIMB

In 2015, an estimated 1.2 million people aged 10–19 died. Many of the leading causes of death, particularly for older adolescents and males, are related to risky behaviours.



SOURCE: WHO

of dopamine receptors, has been linked to the greater sensitivity that teenagers feel to rewards for positive as well as perilous behaviours.

Telzer's studies[7] suggest that teenagers who show heightened ventral striatum activity when making decisions that help others, such as donating money, take fewer risks in the long term and have a lower risk of depression as adults. "There's very much a yin and yang to this," says Dahl.

There are limitations to these lab-based studies; it's hard to reproduce the social whirl of teenage life in a scanner, says Galván. "How do we emulate what's going on on Saturday night in a cold lab on a Tuesday afternoon?" she asks. The studies are more likely to capture a teenager's inclination for risk than the likelihood of real-world risk-taking, Galván says.

The other problem is that the average teenager in a study is only moderately likely to take risks. "Most of what we know about adolescent risk-taking is actually derived from relatively normative samples," says Telzer, "not adolescents engaging in high levels of risk-taking behaviour." Dangerous risk-taking could be confined to a small proportion of teenagers, and there is evidence that they process risk very differently from their peers.

HIGH-RISK RESEARCH

Telzer ran an as-yet-unpublished study in 2015 with adolescents who had been expelled from a school for serious offences. Her team asked them to lie in a scanner and push a button when they saw letters on a screen, but not if the screen displayed an 'X'. Images with social significance — positive pictures such as teenagers laughing or playing games on a beach, and negative ones including a group ganging up on someone — also appeared on screen. Most teenagers were worse at the button-pressing task when the images were positive; their cognitive control was overridden by the rewarding picture. Activity in the ventral striatum went up in tandem. But among the expelled or suspended students, it was the aversive pictures that impaired performance. The teenagers' lack of control, Telzer says, seems to come from a different type of reaction to social stimuli.

Scientists have assumed that the young people who take the most risks show an extreme version of the standard teenager brain profile, says Telzer. But perhaps, she says, they are "a very different type of adolescent".

Research on risk-taking has begun to inform the US justice system. Authorities are taking into account, for example, the factors that might impair a teenager's self-control. Studies show[8] that in emotionally neutral situations, young adults perform cognitive tasks just as well as older adults. But when the situation is emotionally charged, their performance drops off. This and other work could suggest that crimes in emotionally 'cold' situations should be considered differently from those in which 'hot', or emotionally led, decision-making takes over. Similar work could provide ways of pinpointing teenagers at high risk of doing something dangerous.

Steinberg testified in five court cases last year concerning criminal sentences for adolescents. After hearing his evidence on how decision-making in teens is influenced by emotion, a Kentucky court last year decided to raise to 21 the age at which individuals could be given the death penalty. And the evidence has also been enlisted in arguments against mandatory life sentences without parole for offenders under 21.

Scientists are excited about the possibility that this body of developmental research can inform policy. But some, such as Satterthwaite and Galván, point out some challenges in using fMRI data in court for individual cases. The data from neuroimaging studies are usually averaged out across participants, so drawing conclusions about any one brain is itself risky. "Honestly, I don't think neuroimaging should be used," Satterthwaite says. "It's too noisy."

The data are also too noisy for diagnosis, but Satterthwaite is tantalized by evidence that the young brain's response to risk might reveal early symptoms of depression or anxiety. He would like to see research get to the point at which it could guide clinical treatment. "The idea that you can come see me with a life-threatening condition, and leave with no diagnostic test, no imaging, no lab test — that's medieval," he says.

The broader research on adolescent risk is already helping to minimize dangerous behaviour in daily life. For instance, adolescents who don't get enough sleep are more prone to a host of risk-taking behaviours, such as smoking and sexual activity. Dozens of studies[9] on the effects of increasing sleep by delaying school start times — a move endorsed by bodies such as the US Centers for Disease Control and Prevention and the American Academy of Pediatrics — suggest that many of these problems, including risky behaviours, improve when schools start later. The academy recommends a start time of 8:30 or later; hundreds of schools in the United States have delayed their first bell, but in 2014 the median start time for middle school was still 8:00.

Steinberg has advocated limiting exposure to risk in the first place, for example by raising the minimum age for buying tobacco to 21 or prohibiting alcohol sales within 300 metres of schools. This is likely to work better than approaches based on informing students about risks, he says. Other policies aim to take away the opportunity for dangerous behaviour. Graduated-licensing schemes in Australia, New Zealand, Northern Ireland and the United States compel young drivers to build up experience before they are allowed to drive with only teenage passengers. Such programmes have been shown to reduce casualties among young drivers.

But a little bit of risk is a good thing, says Casey. "I wouldn't say that we want people to stop taking risks," she says. "A lot of it is allowing them to be adults in safe situations."

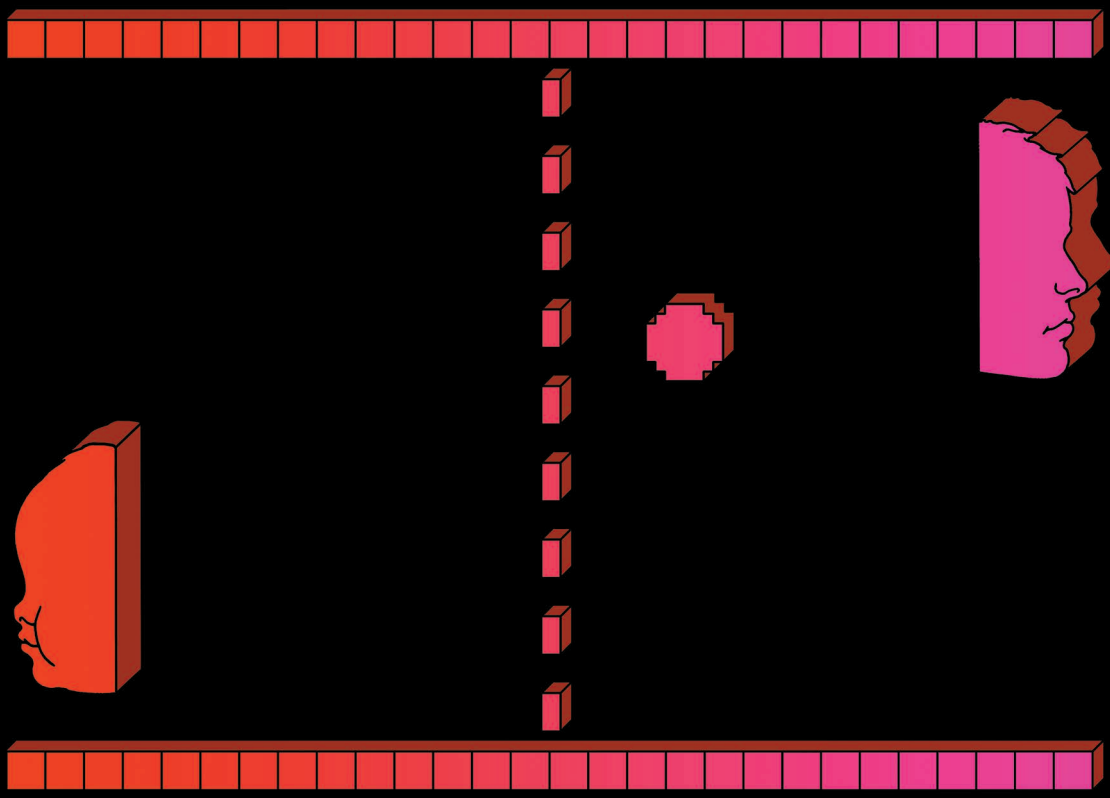
Adolescents have a lot to learn in their transition to relative independence — and nobody said it was easy. "I can't think of a more challenging period of development," says Casey. "Every time I give a talk, I ask people to raise their hand if they want to go through adolescence again. And no one does." ■

Kerri Smith *is a News Feature editor at* Nature, *based in London.*

1. Johnston, L. D. *et al.* Monitoring the Future: National Survey Results on Drug Use, 1975–2016 — 2016 Overview, Key Findings on Adolescent Drug Use (Univ. Michigan, 2017); available at http://go.nature.com/2ngttbt
2. Chein, J., Albert, D., O'Brien, L., Uckert, K. & Steinberg, L. *Dev. Sci.* **14,** F1–F10 (2011).
3. Telzer, E. H., Ichien, N. T. & Qu, Y. *Soc. Cogn. Affect. Neurosci.* **10,** 1383–1391 (2015).
4. Peake, S. J., Dishion, T. J., Stormshak, E. A., Moore, W. E. & Pfeifer, J. H. *Neuroimage* **82,** 23–34 (2013).
5. Telzer, E. H., Miernicki, M. E. & Rudolph, K. D. *Dev. Psychopathol.* **30,** 13–26 (2018).
6. van Hoorn, J. *et al. Res. Adolescence* **26,** 90–100 (2016).
7. Telzer, E. H., Fuligni, A. J., Lieverman, M. D. & Galván, A. *Proc. Natl Acad. Sci. USA* **111,** 6600–6605 (2014).
8. Cohen, A. O. *et al. Psychol. Sci.* **27,** 549–562 (2016).
9. Wheaton, A. G., Chapman, D. P. & Croft, J. B. *J. Sch. Health* **86,** 363–381 (2016).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

THE SHIFTING BOUNDARIES OF ADOLESCENCE



RESEARCHERS STRUGGLE TO DEFINE THE SPAN OF TIME BETWEEN CHILD AND ADULT.

BY HEIDI LEDFORD

On the days when Roy Laver was summoned to be measured, he knew to expect two things. "You were going to be cold," he says. "And you were going to be starkers — in the nude — with lots of adult people looking at you."

Every few months from 1949 to 1971, dozens of youths, Laver included, were marched to a specially outfitted (but poorly heated) wing of the Harpenden branch of the National Children's Home, near London, UK. The home was a charity-run institution that took in neglected youngsters. Laver, now 80, put up with periodic testing there as part of a ground-breaking study to define adolescence.

Each battery of tests took about 3 hours. There were calipers to assess body fat, photographs and X-rays to record growth, and a physical and dental exam. The boys and girls were pleased to miss an afternoon of school, says Laver, but they came to dread the calipers' pinch on their scrawny frames. "If you didn't have a lot of fat, it bloody

ADOLESCENCE
A *Nature* special issue
nature.com/collections/adolescence

ILLUSTRATION BY ERIC NYQUIST

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

hurt," he says. "It took six or seven tries to find some."

The data from that study are still used today to track a child's progression through growth spurts and puberty towards sexual maturity — a journey that the study's lead investigator, paediatrician James Tanner, used to define a firm, physical beginning and end for adolescence.

But few researchers feel that same certainty today. Adolescence is nestled between two transitions — the end of childhood and the beginning of adulthood — with malleable borders. Epidemiological data suggest a trend towards earlier puberty in some countries, including the United States and China; at the other end of the spectrum, societal changes and a growing awareness of the timeline of brain development are pushing the accepted threshold for adulthood well into the twenties.

Scientists, clinicians and policy-makers now find themselves wrestling with these shifting frontiers and asking anew: when does childhood end, when does adulthood begin, and what exactly happens in the middle?

Without a good definition of what adolescence is, it's difficult to nail down what protections and responsibilities adolescents should have. The debate is taking on more urgency as researchers have realized that adolescence helps to set the trajectory for adult life, and as judges and doctors try to fix boundaries on when a person is competent to make adult decisions. "Clarity regarding our conceptualization of adolescence is not just semantic nit-picking," says Jay Giedd, a neuroscientist at the University of California, San Diego. "It has profound implications for clinical, educational and judicial systems."

The lack of a clear definition could also be choking off funding for research on adolescence, which tends to take a back seat to research on early childhood. "One of the reasons adolescence has suffered in terms of financial support and has generally been put at the back of the queue is because the definition is quite difficult," says Anne-Lise Goddings, a developmental neuroscientist at University College London. "It's hard to give money to something when you can't say what it is."

TURNING POINTS

Generations of researchers have painted adolescence as a span of unremitting hazards on the road to adulthood (see page 426). In 1904, US psychologist G. Stanley Hall wrote an influential two-volume opus on adolescence, which he concluded was between the ages of 14 and 24. Hall, who focused his analysis largely on white boys, promoted the idea that adolescence is a time of upheaval. He blamed the mass media — in the form of cheap fictional pamphlets called 'penny dreadfuls' — as well as 'immoral' activities such as drinking and dancing for leading youths astray.

That narrative of fear still prevails, says Nancy Lesko, a sociologist of education at Columbia University in New York City. Lesko worries that the widespread perception of adolescents as irrational risk-takers could bias everyone from policy-makers to researchers. "My guess is that there are lots of changes going on in adolescence that we're missing because adults, including researchers, are still clinging to these dominant narratives," she says.

Half a century after Hall, Tanner embarked on his Harpenden growth study. It began as an effort to track the effects of improved, post-war nutrition on child development, but would become the most influential quantitative picture of adolescence. Tanner's team meticulously detailed changes in height, weight, fat, breast and testicle sizes, the thickness of pubic hair, and a host of other physiological characteristics.

Tanner broke down his data into a sequence that has been simplified into the 'Tanner stages'. Stage I describes the body before the onset of puberty. In Tanner's study, this occurred at about 11 years old, on average, in girls, with boys following six months later. During Stages II, III and IV, the breasts and testes enlarge. Menarche, a woman's first menstruation, is a relatively late event and typically begins at Stage IV. Stage V marks puberty's completion: for the adolescents in the children's home, this happened at about age 15. The value of the study was in defining the progression of events associated with adolescence; Tanner himself noted that the ages at which they occurred varied too much to be of wide use.

There are questions about how relevant Tanner's data are to today's children, says Celia Roberts, a sociologist at Lancaster University, UK. Some of the participants had arrived at the children's home after enduring years of neglect. During Laver's time at the home, nutritious food was in short supply: most meals leant heavily on potatoes, whereas meat, in the form of Spam, was served about twice a week. Laver took to scraping the cooking pots to scrounge a few extra bites, earning him the nickname 'fatty' — although he says he would have been considered slender by today's standards.

Despite its shortcomings, Tanner's catalogue of puberty has become entrenched. Researchers have suggested other scales, some based on different definitions of maturity, others based on hormone levels or voice changes. None has had Tanner's staying power, says Frank Biro, a paediatrician specializing in adolescent medicine at Cincinnati Children's Hospital in Ohio. Even if the growth curves are a little off, the stages provide a useful point of reference, he says. Biro keeps a signed copy of one of Tanner's books on his shelf and still uses the scale in his research.

In today's adolescents, Tanner's stages may be shifting because the onset of puberty — Stage II — is starting earlier in many populations. Tanner already saw signs of this in his studies. Although food was scarce at the children's home, the residents there were better fed than many children living a century before. He thought that this improved nutrition could account for a reduction in the average age at menarche since the mid-nineteenth century[1].

> "THERE IS NO SUCH THING AS AN AVERAGE ADOLESCENT."

That trend towards earlier puberty has continued, and is most notable in girls. A study[2] of Danish children found that the age of onset of breast development dropped by a full year between 1991 and 2006, to just younger than 10 years. In China, between 1985 and 2010, the age at which a girl gets her first period also dropped by a year[3]. Biro's studies have shown differences between ethnic groups: since the late 1990s, the age at which breast development begins has decreased more among white girls in the United States than among African American girls, who were already developing breasts at a younger age.

The evidence strongly suggests that in many parts of the world, this change is at least partly due to the rise in overweight and obese children, Biro says. Several mechanisms might link obesity to the time of onset of puberty. One is that excess body fat increases oestrogen production, which in turn stimulates the growth spurt and breast development associated with puberty. (For boys, the trends are more complex: one study[4] of US boys found that overweight white and African American boys enter puberty earlier than boys who are underweight, but obese boys tend to enter puberty a little later.)

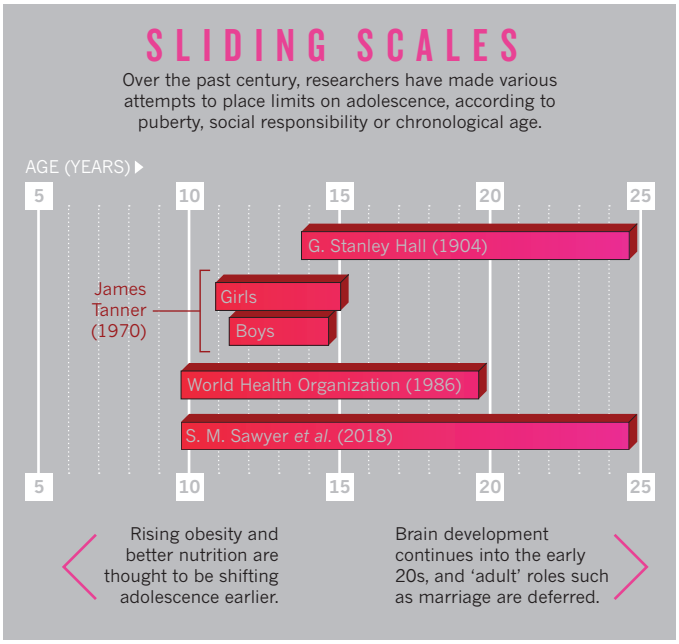
Researchers are trying to decipher what impact this earlier entry into puberty has on the normal progression of adolescence. "You have some of these girls who at age seven and eight are clearly into puberty," says Biro. "But are they adolescents yet?"

In some ways, they might be. Studies have found that earlier-maturing girls exhibit risk-taking behaviours that are typically associated with adolescence[5]. And in a survey of more than 700 magnetic resonance imaging scans from 275 people, Goddings and her colleagues found some changes in the brain that correlated with the Tanner stage of their subjects. For

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

example, the growth of the amygdala, a brain region involved in processing emotions, was influenced by chronological age and by the onset of puberty[6].

Goddings urges caution when interpreting the results: she does not yet know whether the growth patterns trigger behavioural changes common to adolescence or whether they are a response to such changes. A 9-year-old girl, for example, whose breasts have begun to develop might be treated differently by the people around her. This shift could prompt her to behave differently from a prepubescent girl of the same age, with corresponding changes in brain activation. "If a 9-year-old goes through puberty that probably does change something about the brain," says Goddings. "And yet in many ways they are still a 9-year-old."



ON THE THRESHOLD

The definition of the end of adolescence — the emergence into adulthood — is even more nebulous. Although Tanner equated this with the end of puberty, many today use a different definition of adolescence that extends well beyond that point. This could have implications for everything from determining when a criminal can be tried as an adult to when a youth becomes responsible enough to make their own medical decisions.

There is no physical measurement that can capture the beginning of adulthood, say researchers. Often they set the end of adolescence on the basis of social roles. "We don't have an equivalent physical definition of the end of adolescence," says John Coleman, a psychologist at the University of Oxford, UK. "It's lacking a clear definition because it combines social and physical developmental factors."

Those social roles vary widely by culture and era, opening up the definition to a range of interpretations. Tanner tethered the end of puberty to the onset of adulthood, which coincided with social forces in post-war Britain: Laver, for example, left the children's home at 16 to take a job as a clerk in a police station while living with relatives.

But in many societies today, the conventional markers of adulthood are slipping to later in life. Young people spend more years at school, live with their parents for longer, and delay marriage and parenthood. Marriage, in particular, has historically been a key marker for adulthood in many cultures, says anthropologist Alice Schlegel at the University of Arizona in Tucson. The average age of women at first marriage has risen by two years globally over the past two decades, according to the United Nations. In some countries, that increase is more dramatic: in Brazil the average age has increased by 6 years to 27, and in several European countries the age is creeping over 30.

And whereas Tanner could only speculate about changes in the brains of his young subjects, studies by today's neuroscientists have bolstered the idea that adolescence does not stop in the teenage years. The prefrontal cortex, considered the seat of executive functioning and responsible for the ability to plan ahead and to resist impulses, typically does not come fully online until the early twenties — or later[7]. "The brain doesn't suddenly become adult at age 18," says Sarah-Jayne Blakemore, a neuroscientist at University College London.

But if not then, when? The hotchpotch of definitions in research articles, social policies and laws around the globe reveals a wide range of opinion about the end of adolescence (see 'Sliding scales'). The World Health Organization set its boundaries at ages 10 and 19, but Susan Sawyer, chair of adolescent health at the University of Melbourne in Australia, and her colleagues have argued[8] that this upper boundary should be raised to 24. In 2017, New Zealand revised its regulations regarding children in protective care: rather than sending them out on their own at the age of 18, the government continues to provide support into their twenties. The change came in response to reports that the adolescents were not coping well with independence at younger ages.

As helpful as it would be to have a fixed, biological end point to adolescence, neuroscientists are unlikely to derive one, says Blakemore. Ultimately, the end point of adolescence is a social construct, she says, with large differences between cultures. And there is so much variation in brain function and structure from person to person that it may not be possible to nail down a suitable biological end point. "There is no such thing as an average adolescent," Blakemore says.

That can confound research by making it difficult to compare and interpret results across studies. And it complicates social policies by yielding a patchwork of inconsistent guidelines. But Beatriz Luna, who studies neurocognitive development at the University of Pittsburgh in Pennsylvania, sees the differing definitions as a reminder of the complicated and gradual changes that occur in the brain as it specializes for adult modes of action — a process that continues into adulthood. "I don't believe the brain has an abrupt change that will determine the end of adolescence," she says.

At the National Children's Home in the 1950s, however, there was no discussion of when adulthood began, says Laver. At 16, Laver simply left to take a job, later attending a technical college before embarking on a career in the Royal Air Force. During Tanner's study, Laver says he was told that no one would see the pictures that had been taken of him, starkers and shivering, at the National Children's Home. But twice, some of the older students stumbled across them in a nursing magazine. Black bars obscured much of the faces in the photos, but Laver easily recognized some of his classmates.

Many physicians are unaware of the full history of the Tanner scale, says Roberts, who has written about the growth study. "When clinicians have heard me talk, they are quite shocked," she says. "We just wouldn't do that to kids anymore."

But Laver looks back on his days in Harpenden with affection, and with respect for the charity's efforts. The standards of child care were different back then, he says. And participation in the study was voluntary — sort of. "We were told to put your hand up if you would like to have half a day off school," Laver says. "And up went young Roy's hand." ■

Heidi Ledford *reports for* Nature *from London.*

1. Tanner, J. M. *Nutr. Rev.* **39,** 43–55 (1981).
2. Aksglaede, L., Sørensen, K., Petersen, J. H., Skakkebæk, N. E. & Juul, A. *Pediatrics* **123,** e932–e939 (2009).
3. Zhai, L. *et al. PLoS ONE* **10,** e0134656 (2015).
4. Lee, J. M. *et al. Pediatrics* **137,** e20150164 (2016).
5. Johansson, T. & Ritzén, E. M. *Endocr. Dev.* **8,** 126–136 (2005).
6. Goddings, A.-L. *et al. NeuroImage* **88,** 242–251 (2014).
7. Rubia, K. *et al. Neurosci. Biobehav. Rev.* **24,** 13–19 (2000).
8. Sawyer, S. M., Azzopardi, P. S., Wickremathne, D. & Patton, G. C. *Lancet Child Adolesc. Health* **2,** 223–228 (2018).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



DREW ANGERER/GETTY

In Europe, smartphone ownership among young people aged 9 to 16 is 46%, according to a 2014 survey of 7 countries.

Smartphones are bad for some teens, not all

Young people who are already struggling offline might experience greater negative effects of life online, writes **Candice Odgers**.

Last year, I received a phone call from an angry father. He had just read in the newspaper about my research suggesting that some adolescents might benefit from time spent online. Once, he raged, his children had been fully engaged with family and church and had talked non-stop at meal times. Now, as adolescents who were constantly connected to their phones, they had disappeared into their online lives.

He is not alone in his concern. Increasingly, people are claiming that smartphones have destroyed a generation, or that they might be making adolescents lonely and depressed.

After ten years of tracking adolescents' mental health and use of smartphones, I think that such views are misguided. Most young people aged 11–19 (ages vary between studies) are doing well in the digital age. In the United States, a record 84% of students graduated from high school in 2016. Pregnancy, violence, alcohol abuse and smoking have all declined in teenagers in the past 20 years. Similar trends have been observed in other countries[1].

More and better data are crucial. But studies so far do not support fears that digital devices are driving the downfall of a generation. What online activities might be doing, however, is reflecting and even worsening existing vulnerabilities.


ADOLESCENCE
A *Nature* special issue
nature.com/collections/adolescence

SMARTPHONE GENERATION

In the United States, ownership of mobile phones begins early. My colleagues and I surveyed 2,100 children attending public schools in North Carolina in 2015. In that sample, which is likely to be representative of US adolescents, 48% of 11-year-olds told us they owned a mobile phone. Among 14-year-olds, it was 85% (unpublished data; see go.nature.com/2eeffku).

Another survey, done in the same year, indicates that on average, US teens aged 13–18 engage with screen media (from watching television or online videos to reading online and using social media) for more

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

than 6.5 hours each day; mobile devices account for almost half this time[2]. Ownership and usage is also high elsewhere: in a 2014 survey of 9- to 16-year-olds in 7 European countries, 46% owned smartphones[3].

Alongside this increase in the use of digital technology, young people are taking more time to move between childhood and adulthood. Since the 1960s, young people have been delaying social-role transitions such as marriage, childbearing and taking full-time employment[4].

There is also some evidence for an increase in mental-health problems among adolescents. The percentage of US girls aged 12–17 reporting depressive episodes increased by more than 4% between 2004 and 2014, to 17.3%. The proportion of boys doing so in 2014 was 5.7%, a rise of 1.2% since 2004 (ref. 5). Since 1999, the US suicide rate has also increased for every age group, with the most marked rise among adolescent girls[6]. Similar trends among young girls have been observed elsewhere[7].

Various commentators have suggested that young people's rapidly increasing use of digital technologies is accelerating or even driving these behavioural shifts and mental-health trends. In fact, last month, investors released an open letter demanding that technology giant Apple respond to what they see as a "growing body of evidence" detailing the negative consequences of digital devices and social media among young people.

PROS AND CONS

What do the data show?

In the 1990s and early 2000s, US surveys showed that adolescents who reported spending more time online were more likely to also report symptoms of depression and anxiety[8]. But back then, a fraction of adolescents were online — only 14% of the US adult population had access to the Internet in 1995 — and most spent time playing games or talking to strangers in chat rooms. Today, more than 90% of US adolescents are online daily, and much of their time is spent connecting with friends and family whom they share their offline lives with.

A handful of more recent studies, mainly involving university students, not adolescents, have probed for correlations between people's mental health and their use of digital technologies. These have generated a mix of positive, negative and null findings, all with minuscule effect sizes. One of the largest studies so far looked at more than 120,000 UK adolescents in 2017. It found no association between mental well-being and "moderate" use of digital technology, and reported measurable, "albeit small" negative associations for people who had "high levels" of engagement[9]. (Levels were defined according to empirically derived inflection points.)

Meanwhile, a growing body of research conducted over the past decade suggests that time online can actually benefit young people.

A review of 36 studies published between 2002 and 2017 indicates that teens use digital communication to enhance relationships by sharing intimacy, displaying affection and arranging meet-ups and activities[10]. A 2009 longitudinal study of more than 1,300 children and teens also showed that children aged 6–12 who had higher-quality social relationships (defined according to caregivers' descriptions of the children's relationships with friends, caregivers, siblings and teachers) became more-frequent users of e-mail, chats or instant messaging as adolescents aged 12–18. Their offline friendships as adolescents were also more cohesive, as judged by their own descriptions[11].

"Adolescents who encounter more adversity offline seem most likely to experience the negative effects of using smartphones."

Experimental studies, in which subjects play computer games in the lab, have shown that virtual communication (texting a peer they didn't previously know, say) can help adolescents to 'bounce back' after social rejection[12] — such as being excluded from a game with multiple players.

What the data also suggest, however, is that young people from different socio-economic backgrounds are having very different experiences online.

US teens aged 13–18 from families whose total income is less than US$35,000 per year spend, on average, around 4 hours a day watching television and online videos. That's around twice as much time as that spent by their peers from households that have incomes of more than $100,000 per year[2]. In total, low-income teens spend about three hours more each day engaging with screens.

The 2014 study of 3,500 children aged 9–16 from 7 countries in Europe showed that parents in wealthier homes are more likely to "actively mediate" what their child does online. This might be by talking about it, suggesting ways to use the Internet more safely, or joining in and playing computer games, viewing videos or posting alongside their children[3].

In general, the adolescents who encounter more adversity in their offline lives seem most likely to experience the negative effects of using smartphones and other digital devices.

In our 2015 North Carolina survey, teens from low-income families were more likely than more-affluent peers to report that their experiences on social media resulted in offline physical fights, face-to-face confrontations, or them getting into trouble at school (see 'Social-media spillover'). Adolescents who have a history of victimization are more likely to be bullied, solicited and victimized online[13]. Those with behavioural problems, such as difficulties concentrating in class, or a propensity to get into fights, tend to experience more problems on days when they use digital technology more heavily[8].

Other studies conducted over the past decade indicate that adolescents struggling in their offline lives are more likely to have negative online experiences[14]. For example, already-vulnerable young people are more likely to receive negative feedback on social media, experience difficulties regulating their use of the Internet and spend more time 'lurking' — passively viewing others online, rather than actively engaging with them[15].

The 'digital divide' has conventionally referred to differential access to new technologies. That gap still exists, but is shrinking in many countries[16]. In our 2015 survey, 92% of adolescents aged 10–15 from economically disadvantaged homes had access to the Internet, compared with 97% of other teens of the same age. And 65% of those from disadvantaged homes owned a mobile device, compared with 69% of their peers.

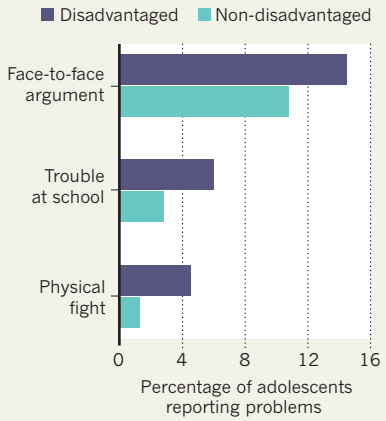
What we're seeing now might be the emergence of a new kind of digital divide, in which differences in online experiences are amplifying risks among already-vulnerable adolescents.

EXPLORE INEQUALITIES

Some might counter that digital technologies are simply providing a fresh medium for the expression of existing problems. They could be right. But given the patterns emerging, it is crucial to investigate thoroughly whether and how the online experiences of adolescents worsen existing inequalities. We ▶

SOCIAL-MEDIA SPILLOVER

In an unpublished survey of 2,100 US teens, those from low-income families are more likely than their affluent peers to report offline problems stemming from use of social media.



SOURCE: C. ODGERS *ET AL.*, UNPUBLISHED DATA

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

must also invest in evidence-based ways to ensure that online experiences are positive for all young people.

This will require advances on several fronts, including the design of rigorous experimental studies. These are challenging because of the difficulty of obtaining control groups — adolescents who are offline or who are willing to have their phones taken away.

One possibility is for researchers to home in on the transition period — when young people first begin to have regular access to mobile devices and social media. Indeed, mobile devices are hugely enabling when conducting research and randomized control trials focused on behaviour and mental health in young people.

Mental states can be gleaned directly from reported information, or indirectly — from data on sleep patterns collected by a wearable device, from entries on Facebook or Twitter, or even from how people text.

Computer scientists, for instance, have predicted the onset of depression from social-media posts and engagement patterns[17]. Also, mobile technologies can be used to deliver 'just-in-time' interventions and support. A 2016 meta-analysis found that brief interventions, such as computer-assisted cognitive behavioural therapy delivered through mobile devices, improved people's psychological well-being and reduced reported symptoms of depression and anxiety[18].

Experimental rigour demands common research protocols, such as standardized questionnaires for assessing online usage and experiences across multiple contexts. The Global Kids Online research toolkit is an excellent example. But such protocols need to be made available in a way that would allow investigators to update them continually, to capture adolescents' evolving digital habits and environments.

The data obtained so far call for other changes, too. Neuroscientists, psychologists and paediatricians need to join forces with those working on human–computer interactions. The Jacobs Foundation 2015 conference on Technologies for Research and Intervention with Children and Youth, at Marbach Castle in Germany, concentrated on building these types of interdisciplinary partnership. Many more such opportunities are needed.

Until a stronger evidence base is built, those who care about the healthy development of adolescents must keep questioning powerful narratives about the next generation. These can blind parents, educators and others to the potential benefits of new technologies for this age group, or, worse, cause the real determinants of mental health and other problems to be missed.

A 2017 petition published in *The Guardian* newspaper called for policies based on evidence, not fear, and was signed by more than 80 scientists (myself included). It offered some pushback against a predominantly one-sided conversation in the media. More crucial is informed and evidence-based dialogue between educators, health professionals, parents, researchers and adolescents.

FAMILIAR TERRITORY

Because online problems can be largely predicted by young people's vulnerabilities offline, much of our existing knowledge about what promotes healthy child development is applicable even in what seems like a foreign digital landscape. Strategies such as the maintenance of supportive parent–child relationships that encourage disclosure, parental involvement in the activities of their children, and the avoidance of overly restrictive or coercive monitoring will help to support adolescents and keep them safe online, just as they do offline.

"Partnerships are key to ensuring that young people, including the most vulnerable, have equal opportunities online."

Leading professional organizations, such as the European Association for Research on Adolescence, the World Economic Forum and the Society for Research in Child Development, could provide important leadership in this regard. Finally, partnerships between local governments, technology companies and educational institutions are key to ensuring that young people, including the most vulnerable, have equal opportunities online.

Social-media sites offer basic protections for adolescent users by providing information to caregivers. But most safety protocols rely on parental advocacy and active mediation and management of online activities, which might leave the most vulnerable young people unprotected.

In December 2017, Facebook pledged $1 million in research funds to help better understand the "relationship between media technologies, youth development and well-being". The best use of such funding could be the development of tools, screening algorithms and outreach strategies for the most vulnerable adolescents. For instance, machine learning and clinical expertise could be leveraged to build classifiers that predict current and future mental-health problems, and such screening algorithms could be used alongside 'just-in-time' interventions.

As Facebook is learning from its recent application of artificial-intelligence approaches to screening for suicide risk, this is not a simple problem. But it is a challenge that technology companies, computer scientists and psychologists are well positioned to take on.

Adults worry about how adolescents spend their time. The telephone, rock 'n' roll, comic books and romance novels all elicited panic. As a parent, I am sympathetic. One in three Internet users worldwide is a child, and the explosion of algorithmically selected content in particular raises legitimate concerns about responsibility and agency.

Yet the design of a digital world that is safe, inclusive, stimulating and nurturing for all requires that we resist fear-based reactions. Instead, we must use the data to understand the very different experiences that young people from diverse backgrounds are having online. ■

Candice Odgers *is a professor of psychology and social behaviour at the University of California, Irvine; a research professor in the Sanford School of Public Policy at Duke University, Durham, North Carolina, USA; and a Fellow of the Jacobs Foundation and the Canadian Institute for Advanced Research.*
e-mail: codgers@uci.edu

1. Sedgh, G., Finer, L. B., Bankole, A., Eilers, M. A. & Singh, S. *J. Adolesc. Health* **56,** 223–230 (2015).
2. Rideout, V. *The Common Sense Census: Media Use by Tweens and Teens* (Common Sense Media, 2015); available at http://go.nature.com/2c4dwew
3. Mascheroni, G. & Ólafsson, K. *Net children Go Mobile: Risks and Opportunities* 2nd edn (Educatt, 2014).
4. Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D. & Patton, G. C. *Lancet Child Adol. Health* http://doi.org/ckdg (2018).
5. Mojtabai, R., Olfson, M. & Han, B. *Pediatrics* **138,** e20161878 (2016).
6. Curtin S.C., Warner M. & Hedegaard, H. *Increase in Suicide in the United States, 1999–2014* NCHS data brief, no. 241 (National Center for Health Statistics, 2016).
7. Collishaw, S. *J. Child Psychol. Psych.* **56,** 370–393 (2015).
8. George, M. J., Russell, M. A., Piontak, J. R. & Odgers, C. L. *Child Dev.* **89,** 78–88 (2018).
9. Przybylski, A. K. & Weinstein, N. *Psychol. Sci.* **28,** 204–215 (2017).
10. Yau, J. C. & Reich, S. M. *Adolesc. Res. Rev.* http://dx.doi.org/10.1007/s40894-017-0059-y (2017).
11. Lee, S. J. *J. Comput.-Mediat. Commun.* **14,** 509–531 (2009).
12. Gross, E. F. *Dev. Psychol.* **45,** 1787–1793 (2009).
13. Kowalski, R. M., Giumetti, G. W., Schroeder, A. N. & Lattanner, M. R. *Psychol. Bull.* **140,** 1073–1137 (2014).
14. George, M. J. & Odgers, C. L. *Persp. Psychol. Sci.* **10,** 832–851 (2015).
15. Underwood, M. K. & Ehrenreich, S. E. *Am. Psychol.* **72,** 144–158 (2017).
16. OECD. *Are There Differences in How Advantaged and Disadvantaged Students Use the Internet? PISA in Focus 64* (OECD Publishing, 2016).
17. De Choudhury, M., Gamon, M., Counts, S. & Horvitz, E. *Proc. 7th Int. Conf. Weblogs Soc. Media* 128–137 (2013).
18. Versluis, A., Verkuil, B., Spinhoven, P., van der Ploeg, M. M. & Brosschot, J. F. *J. Med. Internet Res.* **18,** e152 (2016).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

MOHAMMAD PONIR HOSSAIN/NURPHOTO/GETTY



Children working in a shipyard in Bangladesh. Poor adolescents in low- and middle-income countries often have to work, sometimes in dangerous conditions.

Understand young people in low-income countries

For most of the world's adolescents, poverty and social marginalization influence health much more than risk-taking does, argue **Robert Blum** and **Jo Boyden**.

Nearly 90% of current evidence about adolescence comes from research in high-income countries. Yet nine-tenths of people aged between 10 and 24 live in low- and middle-income countries (LMICs), where this life stage looks very different (see 'Worlds apart').

In LMICs, young people's health and well-being tend to be more severely affected by cultural, socio-economic and environmental risk factors than in high-income countries, and there are fewer resources to mitigate such risks. Adolescents in LMICs also tend to have many more family responsibilities than their peers in high-income countries.

Currently, millions of young people in LMICs are condemned to poor health, impaired development and premature death. In fact, since 1990, there have been fewer improvements in health for adolescents (aged 10–19) and young people (aged 15–24) in these countries than for any other age group[1].

We need to better understand the everyday realities of adolescents' lives in LMICs and how this affects their health. This would enable targeted investment to improve well-being and productivity globally — but it

ADOLESCENCE
A *Nature* special issue
nature.com/collections/adolescence

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

means abandoning Western assumptions about adolescence being predominantly a time for risk-taking. It also means finding better ways to reach adolescents — for instance, through the use of mobile phones, social media and community-based strategies — and applying approaches such as web-based surveys and interactive data-collection methods to build a richer picture of the forces shaping young people's lives.

POVERTY TRAP

Studies show that adolescents in LMICs who are subject to poverty and restricted access to health, education and other services are also more likely to be exposed to environmental toxins and extreme weather events, such as droughts, than their wealthier peers.

And this pile-up of multiple stressors is likely to worsen in the coming years. According to the Notre-Dame Global Adaptation Initiative (http://gain.nd.edu), out of the 100 countries most vulnerable to climate change, 42 are low-income countries, 33 are lower middle-income countries and 14 are upper middle-income countries. Meanwhile, rapid and unplanned urbanization is driving more overcrowding, road injuries, noise pollution and the accumulation of toxins such as lead[2].

In LMICs, adolescents from minority communities tend to be significantly more disadvantaged than those in majority populations. A 2009 study of the health and well-being of more than 10,000 individuals in Vietnam aged 14 to 25 found that young people from ethnic minorities fared worse than those from the Kinh majority in every measure[3]. For example, 10% of young people from ethnic minorities reported being illiterate, compared with 1% of Kinh young people.

Gender is another major determinant of health differences. The pilot of the ongoing Global Early Adolescent Study (www.geastudy.org/), which R.B. is directing, collected predominantly quantitative data between 2015 and 2017 from about 35 families in 15 countries on 5 continents. These data indicate that, compared with their brothers and male peers, girls in LMICs tend to experience more social isolation as they move from childhood to adolescence, and have fewer opportunities in education, recreation and exploration.

Opportunities for education can be just as important as nutrition for adolescent health. According to a 2014 study, for instance, there is a strong link between teenage pregnancy and low levels of literacy in Africa[4]. And a wealth of literature shows that deficiencies early in life cast a long shadow.

As part of Young Lives (www.younglives.org.uk), an ongoing longitudinal study directed by J.B., researchers have been collecting information about 12,000 children in Ethiopia, India, Peru and Vietnam for 15 years. In all four countries, children whose growth is stunted when they are around one year old are likely to remain so until they are at least 15 (ref. 5). Those who experience an early nutritional disadvantage are more likely to have difficulty progressing through school. And stunting at age 8 correlates with lower scores on measures of self-efficacy, self-esteem and educational aspirations at age 12 (ref. 6).

"Adolescents are the least likely of any age group to access conventional health services."

Another characteristic of life for poor adolescents in LMICs is combining the pursuit of education with work. Young Lives data suggest that the time devoted to education remains broadly constant for girls and boys aged 8 to 15. But as they grow older, adolescents tend to do more work, either for pay or as part of family life. Indeed, they are frequently the primary carers of younger siblings or incapacitated adults. Some are even the principal breadwinners in their households (see 'Age of responsibility').

Employment can affect health during adolescence in many ways. Young workers commonly suffer more accidents than adult workers, and they are particularly susceptible when exposed to chemical toxins from mining or agricultural work, for example[7].

NO TIME TO PLAY

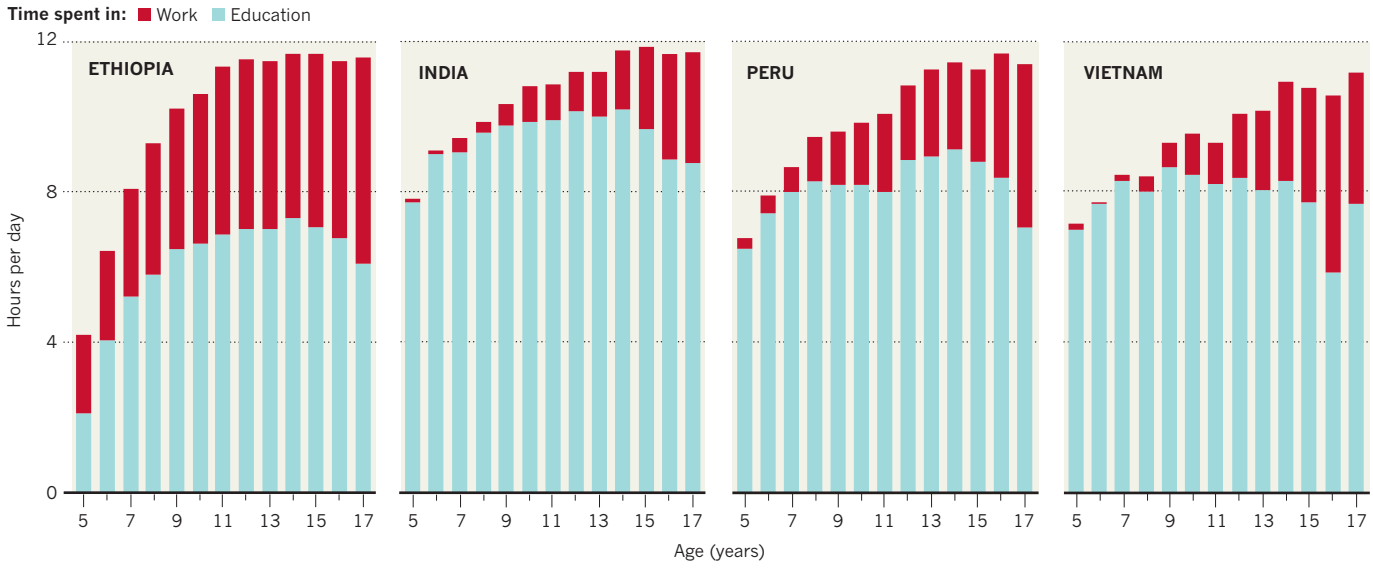
Experts in the Global North generally assume that adolescence is a carefree time of emerging independence, social exploration and risk-taking. This way of thinking carries over to adolescent-focused health programmes and policies in LMICs, which regularly focus on violence, sexually transmitted infections and teen pregnancies.

Too many programmes fail to understand what drives behaviour in the first place, and ignore the broader risks that young people face from poverty, work, social stigma or exclusion from quality services. Also, many centre around the provision of clinical services, even though adolescents are the least likely of any age group to access conventional health services[8].

With the right approach, community-based interventions geared towards reducing a behavioural problem such as violence can also improve the overall health and well-being of young people. Reducing violence

AGE OF RESPONSIBILITY

As they grow up, young people in low- and middle-income countries tend to devote more of their time to work that helps sustain their families. Under-reporting of work is likely because those surveyed described 'a typical day last week' when school was in session.



SOURCE: P. ESPINOZA REVOLLO & C. PORTER, UNPUBLISHED DATA

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



An apple seller in Changzhi, China. Text messages can provide health information for adolescents.

REUTERS

and the incidence of HIV/AIDS are among the principal aims of South Africa's activity-based Waves for Change surf schools, Soul Buddyz Clubs and AMANDLA EduFootball programmes, for example. But these programmes also improve young people's fitness, psychological well-being and ability to form friendships and take responsibility for others.

Donors, governments and non-governmental organizations in LMICs are expanding their repertoire of approaches. But better research tools and more community-based initiatives are needed that focus specifically on understanding and enhancing resilience among adolescents.

Digital technologies offer fresh ways to improve health and well-being (see page 432). Various surveys conducted over the past decade indicate that young people in LMICs value the unrestricted access to information and the privacy that the Internet affords, and prefer digital media over other delivery channels for health information[9]. Several investigators have already tried to exploit these preferences. In a recent randomized trial in Ghana, for instance, text messages improved girls' knowledge about reproductive health[10]. Evidence on the effectiveness of digital health information in changing behaviours is needed, however.

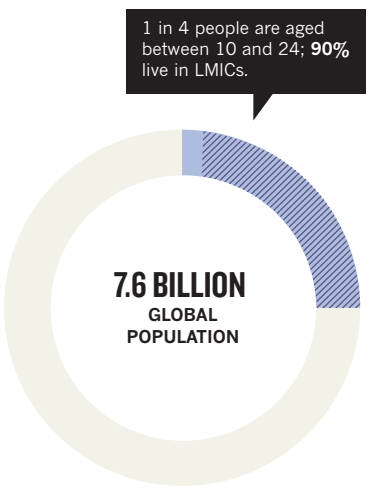
Importantly, researchers, policymakers, practitioners and others must deploy a bottom-up approach in which adolescents are treated as partners in improving their health and well-being, not just the recipients of change. Likewise, the design and implementation of interventions must draw on evidence of what works, for whom and why.

This means that investigators must work in communities and across all settings — not just in hospitals and clinics — to understand how young people spend their time, how they access resources, who controls their earnings, and so on. Researchers must also work out how health measures can be integrated into schools or mobile-phone applications, rather than delivered solely through medical facilities. This is not simply about making schools a hub for health education and services; greater flexibility in timetables or the provision of catch-up classes and vocational measures, for example, could encourage more working adolescents to stay in school for longer.

Finally, international organizations, such as the World Health Organization, the International Labour Organization and the United Nations Children's Fund, should direct much greater effort towards economic analyses and surveys of occupational and environmental hazards, exposures to social risks and the mental health of young people in LMICs.

The second decade of life presents an extraordinary opportunity to improve people's health over the long term. And threats to human health and well-being are becoming more acute in our rapidly urbanizing, industrializing world, with rising environmental risks and the potential for reductions in youth employment opportunities as a result of expanding technologies.

Only with a realistic understanding of the lives of young people in LMICs, grounded in the social, economic and political contexts of their everyday lives, do we stand a chance of shaping their futures for the better. ■



SOURCE: A. KÅGESTEN *ET AL. PLOS ONE* **11,** E0157805 (2016)

Robert Blum *is the director of the Johns Hopkins Urban Health Institute in Baltimore, Maryland, USA, and lead researcher on the Global Early Adolescent Study.* **Jo Boyden** *is professor of international development and director of the Young Lives study at the University of Oxford, UK.*
e-mails: rblum@jhu.edu; jo.boyden@qeh.ox.ac.uk

1. Mokdad, A. H. *et al. Lancet* **387,** 2383–2401 (2016).
2. Wachs, T. D. & Rahman, A. in *Handbook of Early Childhood Development Research and Its Impact on Global Policy* (eds Britto, P. R., Engle, P. L. & Super, C. M.) 85–122 (Oxford Univ. Press, 2013).
3. Vietnamese Ministry of Health & General Statistics Office of Vietnam. *Survey and Assessment of Vietnamese Youth* (2009).
4. Odejimi, O. & Bellingham-Young, D. *J. Hum. Growth Dev.* **24,** 135–141 (2014).
5. Georgiadis, A. *et al. SSM Popul. Health* **2,** 43–54 (2016).
6. Dercon, S. & Sánchez, A. *Econ. Hum. Biol.* **11,** 426–432 (2013).
7. International Programme on the Elimination of Child Labour. *Children in Hazardous Work* (ILO, 2011).
8. Patton, G. C. *et al. Lancet* **387,** 2423–2478 (2016).
9. Byrne, J., Kardefelt-Winther, D., Livingstone, S. & Stoilova, M. *Global Kids Online Research Synthesis, 2015–2016* (UNICEF Office of Research Innocenti and London School of Economics and Political Science, 2016).
10. Rokicki, S., Cohen, J., Salomon, J. A. & Fink, G. *Am. J. Publ. Health* **107,** 298–305 (2017).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

In Retrospect

Pruning hypothesis comes of age

The idea that disrupted pruning of neuronal connections in the brain during adolescence is a cause of schizophrenia was proposed in 1983. This proved prescient, as subsequent imaging, genetic and molecular research has shown.

MATTHEW B. JOHNSON & BETH STEVENS

It is common knowledge among neuroscientists today that the brain's frontal lobe, which is involved in many complex human behaviours, continues to develop throughout adolescence and into early adulthood. But in 1983, the idea of such protracted brain development was, in the words of the psychiatrist Irwin Feinberg, "a newly emerging theme"[1]. That year, he published the hypothesis[1] that schizophrenia is caused by defects in a particular aspect of adolescent brain development. Genomic studies are now beginning to shed long-awaited light on the mechanisms that might underlie such a process.

Among the earliest contributions to this emerging field was a landmark 1979 study by the neuroscientist Peter Huttenlocher. Using electron microscopy, Huttenlocher imaged and counted the connections, called synapses, between neurons in slices of the frontal lobe from individuals ranging in age from newborn to 90 years old. Unexpectedly, he found that a drastic reduction occurred in the number of synapses between infancy and adulthood[2]. At the time, Feinberg was studying sleep patterns during human development and ageing. The changes detailed in Huttenlocher's report were strikingly similar to a phenomenon Feinberg had observed to occur during the second decade of life — a steep reduction in the amplitude of waves of brain activity during dreamless sleep, as measured using a technique called electroencephalography (EEG)[3].

Taking this information together with the propensity for symptoms of schizophrenia to emerge in patients' late teens and early twenties[4], Feinberg boldly speculated that the three phenomena might be linked. According to his hypothesis, normal changes in synapse numbers or organization, determined by genetic programs of synaptic pruning in the adolescent brain, could reduce the amplitude of sleep EEG waveforms. Disturbances of these programs (causing either too much or too little pruning) might produce or contribute to the symptoms of schizophrenia.

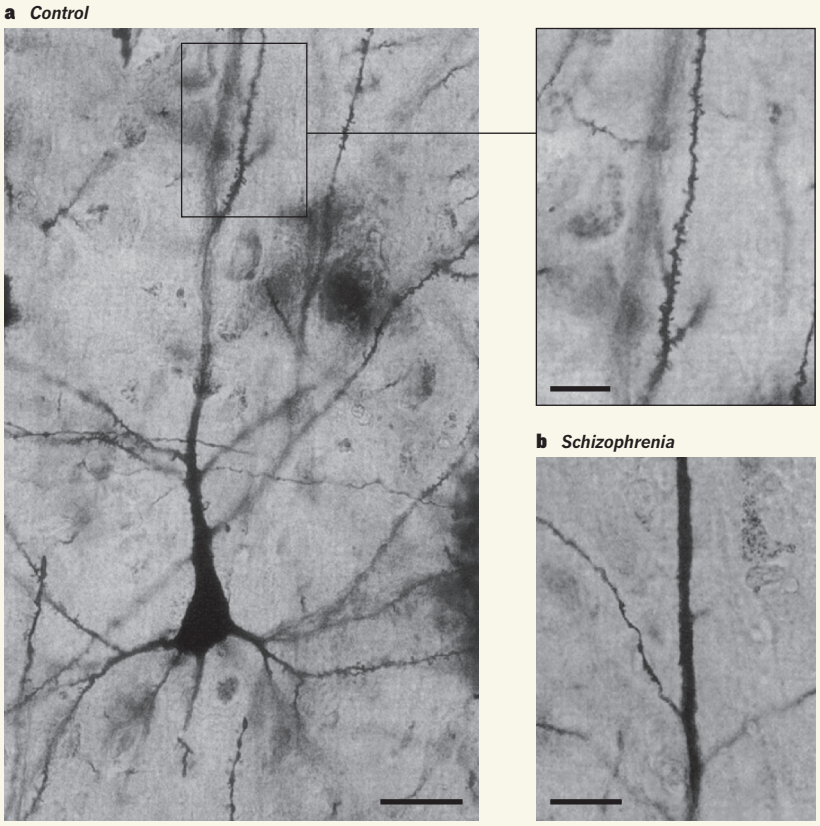


Figure 1 | Photomicrographs of neurons in the human brain. **a**, A neuron from a healthy brain is dotted with structures called dendritic spines (seen in the zoomed-in image). Spine numbers can be used as a proxy for the number of synaptic connections a neuron receives from other neurons (not visible). Scale bar in main image, 30 μm; in zoomed-in image, 20 μm. **b**, There are fewer spines, and so synapses, in the brain of a person who has schizophrenia. Scale bar, 15 μm. (Figure adapted from ref. 8.)

Feinberg had been considering the relationship between brain structure and function since the early 1960s, and recalls sitting in the office of his then mentor, Jean Piaget, a pioneer in the field of cognitive development, and "raising the possibility that maybe part of [cognitive] development was based on biological changes in the brain. He went on with the conversation as though I had not said anything". Twenty years later, neuroscientists reacted similarly to his hypothesis on pruning and schizophrenia, with Feinberg describing the initial response in an interview last year as "totally indifferent" (see go.nature.com/2cohqjx).

Indeed, a major hole in Huttenlocher's analysis remained a sticking point for years. Owing to the difficulty of obtaining samples, and the painstaking analysis involved, his study included only one adolescent. Although data from primates revealed similar postnatal synaptic pruning (for a review, see ref. 5), it was not until 1997 that additional human data supporting this phenomenon became available from a handful of adolescent brains[6]. In 2011,


ADOLESCENCE
A *Nature* special issue
nature.com/collections/adolescence

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

5 more samples from individuals between 10 and 22 years old were finally added to the literature[7].

As researchers were working to fill this gap in the data on normal development, at least two groups were looking for — and, at the turn of the twenty-first century, found — signs of aberrant pruning in the brains of people with schizophrenia[8,9] (Fig. 1). Using a staining technique that enabled them to count the number of neuronal structures called dendritic spines that receive certain synaptic inputs, these researchers provided evidence that synapse densities in brain regions involved in higher cognition are lower in affected than in unaffected people.

Subsequently, the development of non-invasive brain-imaging techniques made the analysis of both healthy brain development and psychiatric disorders much easier. One imaging study[10] confirmed that the volume of grey matter (the brain tissue that generally contains synapses) in the temporal and frontal lobes increases during early childhood and then decreases over the course of adolescence. These changes are consistent with the pruning of neuronal processes and synapses in adolescence. Another study[11] revealed that at-risk individuals who go on to develop psychosis exhibit an accelerated rate of grey-matter loss in the frontal lobe compared with those who do not. Evidence to support Feinberg's hypothesis was gradually growing.

More recently, the hypothesis has been strengthened, thanks to the genomics revolution. Studies[12,13] of tens of thousands of people with schizophrenia have identified genetic variants associated with an increased risk of the disease. By far the strongest signal has been found in a portion of chromosome 6 that contains hundreds of genes related to immunity, making it hard to pinpoint individual risk genes. But in 2016, analyses of brain and DNA samples from affected and unaffected individuals enabled geneticists to pick out variation in one gene in the region, that encoding complement component 4 (C4), as partly responsible for schizophrenia risk[14].

This work revealed that different variants of the *C4* gene produce different levels of the protein it encodes, and that higher *C4* expression is correlated with a higher risk of schizophrenia. Independently of this work, our group and others have shown that the immune signalling cascade of which the C4 protein is a part, called the complement cascade, instructs the brain's immune cells to prune synapses during development in mice (Fig. 2)[15,16]. How this pruning mechanism relates to schizophrenia risk or age of onset, or to normal adolescent changes in brain activity during sleep, is an exciting and active area of investigation. Precisely how and where in the human brain C4 is targeted to synapses, whether particular synapses or circuits are more or less vulnerable to complement-mediated pruning, and which of these mechanisms are most relevant to specific symptoms of schizophrenia, remain unknown.

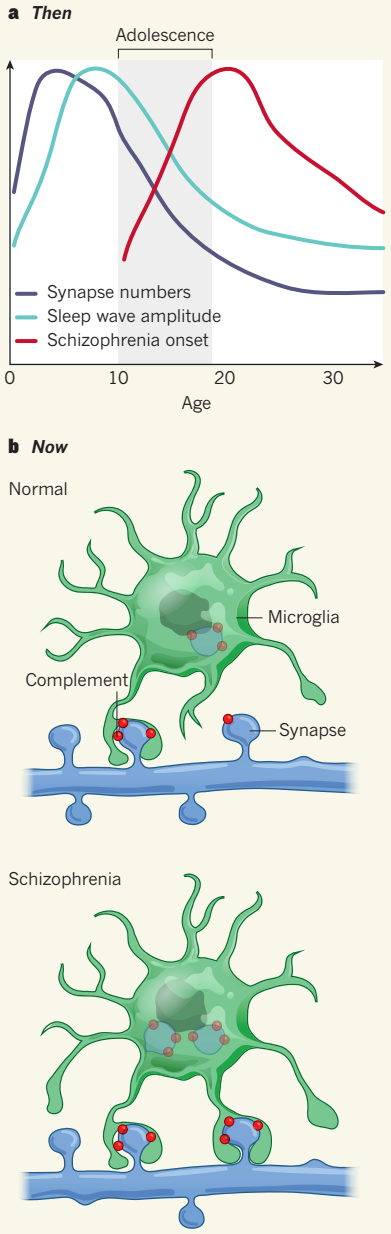


Figure 2 | A growing understanding of brain rewiring in adolescence. **a**, In 1979, an analysis of synapse numbers in the frontal lobe at different ages revealed a dramatic decrease in synapses during adolescence[2]. Sleep brainwave amplitudes follow a similar pattern[3]. Furthermore, symptoms of schizophrenia typically emerge in adolescence[4]. In 1983, these facts together led the psychiatrist Irwin Feinberg, who was studying sleep, to propose[1] that defects in adolescent pruning of synapses might be a cause of schizophrenia. **b**, In the past decade, molecular and genetic studies[12–16] have provided evidence that this pruning defect involves genes of the complement signalling cascade. During normal development, brain-resident immune cells called microglia engulf and prune synapses that are tagged by complement proteins. Higher expression of one complement gene, *C4*, is associated with a higher risk of schizophrenia, perhaps because of higher pruning levels.

Other gene variants associated with schizophrenia risk are scattered throughout the genome, implicating a variety of other synaptic proteins[13]. Thus, synaptic pruning is likely to be disrupted through different mechanisms in different patients. New tools are now enabling us to better understand this variability. For instance, methods for converting human stem or skin cells into brain cells and 3D 'mini-brains' allow us to probe synapse development and pruning using living patients' own cells, or cells engineered to harbour risk variants. In addition, genome-editing tools such as CRISPR–Cas could be used to generate new primate models of frontal-lobe synapse development and psychiatric disorders.

One could charitably interpret the initial resistance to Feinberg's pruning hypothesis as tacit recognition that, at that time, such hypotheses could only be pure, untestable speculation. Indeed, Feinberg himself wrote, "The probability of guessing correctly the brain mechanisms causing schizophrenia is, with our limited knowledge, vanishingly small"[1]. And yet, although we are just beginning to understand the molecular processes underlying his hypothesis, evidence is now accumulating that Feinberg might — at least in part — have done just that. ■

Matthew B. Johnson *and* **Beth Stevens** *are at the F. M. Kirby Neurobiology Center, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA, and at the Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts.*
e-mails: johnson@broadinstitute.org; beth.stevens@childrens.harvard.edu

1. Feinberg, I. *J. Psychiatr. Res.* **17,** 319–334 (1982–83).
2. Huttenlocher, P. R. *Brain Res.* **163,** 195–205 (1979).
3. Feinberg, I., Koresko, R. L. & Heller, N. *J. Psychiatr. Res.* **5,** 107–144 (1967).
4. Sham, P. C., MacLean, C. J. & Kendler, K. S. *Acta Psychiatr. Scand.* **89,** 135–141 (1994).
5. Elston, G. N. & Fujita, I. *Front. Neuroanat.* **8,** 78 (2014).
6. Huttenlocher, P. R. & Dabholkar, A. S. *J. Comp. Neurol.* **387,** 167–178 (1997).
7. Petanjek, Z. *et al. Proc. Natl Acad. Sci. USA* **108,** 13281–13286 (2011).
8. Garey, L. J. *et al. J. Neurol. Neurosurg. Psychiatry* **65,** 446–453 (1998).
9. Glantz, L. A. & Lewis, D. A. *Arch. Gen. Psychiatry* **57,** 65–73 (2000).
10. Gogtay, N. *et al. Proc. Natl Acad. Sci. USA* **101,** 8174–8179 (2004).
11. Cannon, T. D. *et al. Biol. Psychiatry* **77,** 147–157 (2015).
12. The International Schizophrenia Consortium. *Nature* **460,** 748–752 (2009).
13. Schizophrenia Working Group of the Psychiatric Genomics Consortium. *Nature* **511,** 421–427 (2014).
14. Sekar, A. *et al. Nature* **530,** 177–183 (2016).
15. Stevens, B. *et al. Cell* **131,** 1164–1178 (2007).
16. Schafer, D. P. *et al. Neuron* **74,** 691–705 (2012).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

PERSPECTIVE

doi:10.1038/nature25770

Importance of investing in adolescence from a developmental science perspective

Ronald E. Dahl[1], Nicholas B. Allen[2], Linda Wilbrecht[3] & Ahna Ballonoff Suleiman[4]

This review summarizes the case for investing in adolescence as a period of rapid growth, learning, adaptation, and formational neurobiological development. Adolescence is a dynamic maturational period during which young lives can pivot rapidly—in both negative and positive directions. Scientific progress in understanding adolescent development provides actionable insights into windows of opportunity during which policies can have a positive impact on developmental trajectories relating to health, education, and social and economic success. Given current global changes and challenges that affect adolescents, there is a compelling need to leverage these advances in developmental science to inform strategic investments in adolescent health.

It is well-recognized that the first few years of life represent a sensitive period for growth, early learning and brain maturation[1]. As a result, the scientific understanding of the formational effects of early experience has had a pronounced impact on global policies and practices[2,3]. Recent scientific advances are now contributing to the growing recognition that adolescence is a second period of rapid growth and foundational learning associated with distinct neuro-maturational changes[4,5]. There is great potential for leveraging these advances to inform strategies for investing resources during this pivotal time in the life course[6,7].

Notably, the global stakes for choosing to invest in adolescence are increasing. Global demographics reveal a disproportionate surge in the number of adolescents in the world, including an unprecedented expansion of adolescents in low-income regions. More than a billion adolescents are now coming of age among an increasing array of social changes and global challenges. The revolutionary changes in information technology that are sweeping across the globe are having the greatest impact on adolescents. Adolescents are often the early adopters of new technology in ways that can intensify vulnerabilities (for example, exploitation and radicalization) as well as opportunities (education, social connection, innovation and learning). These accelerating technological and social changes heighten the importance of understanding how (and when) to intervene to prevent harmful effects and to promote health.

In this review, we present a developmental science perspective on adolescence as a distinct maturational period that begins with the onset of puberty. Developmental science is the study of the patterns and processes of biological, cognitive and behavioural changes that occur as an organism grows and matures. Here, we summarize our current understanding of developmental processes that occur during and after puberty. We describe some of the distinctive aspects of learning in adolescence that are thought to support acquisition of the culturally embodied knowledge, skills and self-regulatory capacities that are needed for adolescents to become independent and to integrate into adult society. We explore how a nuanced understanding of the distinctive features of this developmental period may inform intervention and policy, especially the developmental changes in learning and exploration that are amplified during adolescence. Specifically, we highlight the role that pubertal hormones may have in enhancing the motivational importance of status, prestige and sex as notable drivers of social learning and identity development during adolescence. Finally, we discuss the exciting potential of using insights from developmental science to inform intervention and policy—particularly in light of current global challenges.

Puberty initiates a distinct developmental period

Adolescence encompasses the numerous developmental changes and foundational learning experiences that occur during the transition from childhood to the attainment of adulthood. This transition begins with the onset of puberty—a biological process that drives sexual maturation, typically starting by age 10 in girls and by age 12 in boys (Fig. 1). Puberty begins with changes in the brain, which initiate a cascade of transformational changes in the body that include rapid acceleration in physical growth (pubertal growth spurt), metabolic changes, changes in sleep and circadian regulation, and sexual maturation (that is, sex-specific changes in facial structure, voice, muscle and fat distribution, body hair distributions, changes in skin and glandular secretions, and breast, genital, adrenal and gonadal development)[8]. Hormones released from the gonads and adrenals, in turn, affect the brain, altering cognitive, emotional and motivational processes (It is important to note that although puberty is often thought of as a developmental process that is sexually differentiated, testosterone and oestrogens can be found in the brain and blood of both males and females with variable, and sometimes overlapping, concentrations.).

Developmental changes during adolescence also include structural and functional changes in the brain—particularly neural systems involved in cognitive, emotional, social and motivational processes[4,9,10]. These neural changes are associated with behavioural changes such as increases in sensation-seeking and a re-orientation of attention and motivation (towards peers, social evaluation, status and prestige, and sexual and romantic interests)[11,12]. Adolescent development also involves profound changes in social contexts, social roles and social responsibilities[13]. Importantly, there are complex interactions between and among these levels of change. Indeed, learning and brain development are inextricably intertwined throughout this period as learning affects brain development, and maturational changes in the brain in turn affect learning and motivation.

The end point of adolescence is considerably more difficult to define scientifically. Becoming an adult cannot simply be equated to a measure of physical maturation, such as attaining final adult height or reproductive maturity. There have also been some recent attempts to define

[1]School of Public Health, Institute for Human Development, University of California, Berkeley, California, USA. [2]Department of Psychology, University of Oregon, Eugene, Oregon, USA. [3]Department of Psychology, Helen Wills Neuroscience Institute, University of California, Berkeley, California, USA. [4]Institute for Human Development, University of California, Berkeley, California, USA.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

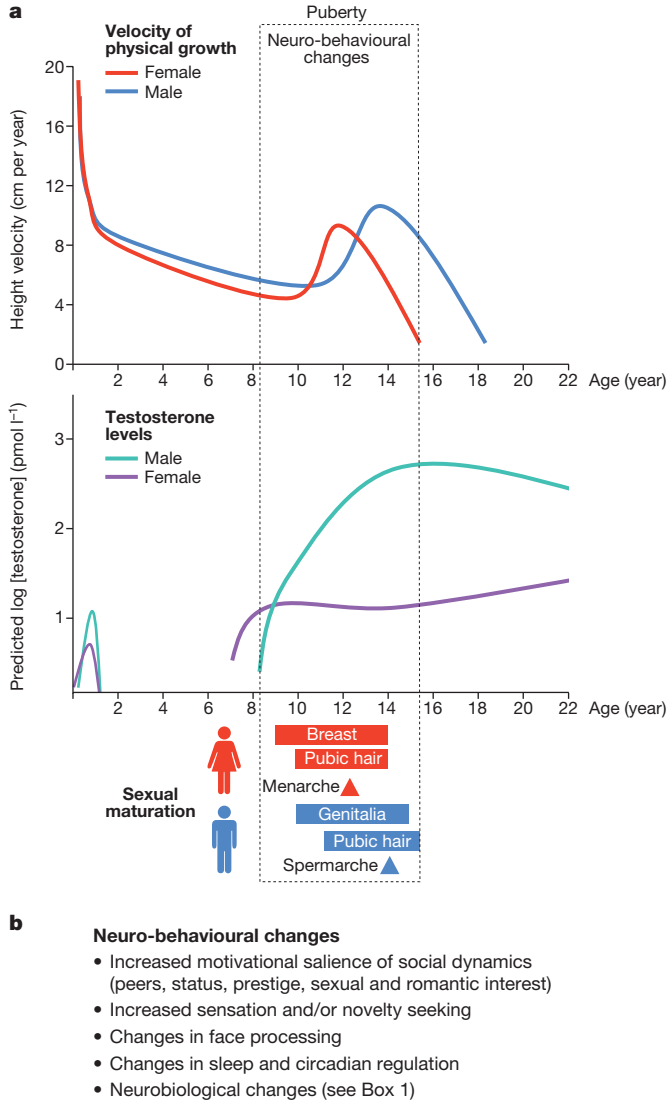


Figure 1 | Puberty initiates a period of rapid growth and multi-level dynamic change. a, The red and blue solid lines (height velocity in cm per year as a function of age in girls and boys) show that growth is very rapid in infancy and then slows during childhood. Growth rates accelerate at the onset of puberty and then decelerate at later stages of puberty. This dynamic pubertal growth spurt begins about 18 months earlier in girls than in boys—and is part of a large collection of physical manifestations of sexual maturation (which in girls includes breast and genital development, pubic and axillary hair and menarche, and in boys includes genital development, pubic, axillary and facial hair). The red and blue lines represent testosterone levels (log(pmol l^{-1})) as an example of one of several hormones that contribute to these pubertal changes. **b**, Other physical changes include sexually dimorphic changes in facial architecture, changes in skin and sebaceous glands (for example, acne and body odour) and deepening of the voice in boys. Pubertal maturation also influences neuro-behavioural development, particularly social, emotional and motivational processes, such as the well-recognized increases in sensation-seeking, and a re-orientation of attention and motivational salience towards peers, social evaluation, status and prestige, and sexual and romantic interests. Pubertal maturation also influences (and is influenced by) the rapidly changing social contexts in which adolescent learning and development are occurring. Taken together these multi-level changes (and their bidirectional interactions) create a dynamic period of growth, development, learning and adaptation. Understanding these dynamic developmental processes can inform targets (and developmental timing) for early intervention, prevention and policy.

the end of adolescence in terms of cognitive and affective processes[14]. However, 'adulthood' is inherently intertwined with taking on certain social roles, and importantly, having the knowledge, skills and social competence to succeed in these roles. It also involves societal recognition that the individual has the rights (and responsibilities) to exercise these roles. As a pragmatic matter, modern societies typically confer adult rights and responsibilities based on age. However, this is at best an approximate way to estimate maturation—especially during adolescence when punctuated growth processes can result in markedly different maturational development between individuals of the same age. Moreover, laws can vary widely (within and across countries) regarding which age is used to designate specific adult rights, roles or responsibilities (for example, the legal age for driving a car, voting in an election, marriage, serving in the armed forces, being sanctioned as an adult for committing crimes, using alcohol, cigarettes or other drugs, making personal decisions about health or owning a gun)[15]. Typically, these legal frameworks are primarily informed by cultural (and sometimes religious) values, rather than evidence about the developmental effects of exposure to particular learning experiences and independent decision making at different stages of development[16].

Learning, development and maturation in adolescence

Adolescent development builds on considerable childhood learning, and allows adolescents to acquire a new level of knowledge, skills and cultural competence to successfully transition to an independent adult role. During adolescence, we gain cognitive, affective and self-regulatory abilities that allow us to adaptively pursue new goals and priorities that can be increasingly abstract and extend far into the future[4,17]. We gain more nuanced understanding of the workings of adult society, which involves learning to reason about abstract concepts and future consequences, and creating a more mature understanding of self, others and the complexities of adult social relationships[17–21]. We also acquire skills to navigate a growing range of novel, uncertain and emotionally charged social situations (including a new level of cognitive, social, emotional and self-regulatory capacities[22]) and to apply these skills to an increasingly complex set of social relationships with peers, adults and societal institutions.

In addition, during adolescence, we must also adapt to fundamentally new aspects of our own emerging identity[23], which includes learning how to relate to the world, and ourselves, as a suddenly and (often) mystifyingly sexual being[24]. We must discover in a new way—through an elaborate series of trials and errors, successes and failures, and with growing independence—who we are. That is, we must develop (within a family, cultural and social context) an adult identity, with our own heartfelt goals, values and priorities.

As supported by a large number of studies, this developmental trajectory from childhood to emerging adulthood is fraught with a multitude of risks and vulnerabilities. These contribute to a marked increase in risk of death and disability through adolescent accidents, suicide, violence, depression, alcohol and substance use, sexually transmitted diseases, unwanted pregnancies, as well as the establishment of a wide range of health-related behavioural risk factors (such as smoking, drinking, substance use, unhealthy eating and sedentary behaviour) that will contribute to health consequences in later life[25]. Pivotal changes in educational trajectories during adolescence have increasingly important lifetime consequences, with educational attainment impacting career options and economic success[15].

The rapidly changing world of adolescence

We are currently witnessing pronounced and historically unprecedented changes in the demography and lifestyle of adolescents[15]. In 2015, the estimated global population of adolescents aged 10–19 years was 1.2 billion, which is approximately 16% of the world's population[26]. The vast majority of the world's adolescents—approximately 90%—live in low- and middle-income countries, in which the barriers to achieving positive health and well-being are often the most complex and challenging. As such, nations that are already being challenged to

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

meet the educational and health needs of adolescents are likely to need increased resources to address these issues in the near future.

Technological innovations, especially in mobile computing (for example, smartphones) and online social networks, are transforming the daily lives of adolescents[27–29]. Today's adolescents in middle- and high-income countries are 'digital natives'—individuals who have never experienced the world before the internet. The social capabilities of internet-enabled devices tap into core adolescent motivations in powerful ways. Adolescents are particularly motivated to explore peer relationships, and social media provide almost ubiquitous access to these interactions[29]. Moreover, it provides these opportunities in ways that are particularly appealing to adolescents—relatively free of the reach of parental monitoring[30]. This creates unprecedented opportunities for positive social connection and support (for example, the isolated teen who finds a community of like-minded individuals online), but it also intensifies adolescents' exposure to negative social encounters (for example, ostracism, bullying, sexual exploitation and violent radicalization). Adolescents' emerging interests in sex and sexuality are now met with an almost unlimited supply of freely available highly explicit pornography. Importantly, these technologies can also add to the 'digital divide', in which social inequalities between those adolescents growing up in high-resource contexts (in which parents, teachers and social influences help to scaffold adolescents' learning and use of technology) and those growing up in impoverished contexts (who may be more vulnerable to the negative effects) are exacerbated[31].

Other factors are also rapidly altering the lives of adolescents. Today, adolescents enter puberty earlier, and take on independent adult roles (for example, finishing formal education and entering the workforce, living apart from their family of origin, becoming fully financially independent) later than they did in the past centuries, making adolescence a longer period of life than it has been previously[32].

Economically, adolescents increasingly take on the role of active consumers, with their own disposable cash and credit accounts making them independently responsible for purchasing decisions[33]. As a result, adolescents are now the direct target of marketing efforts (including marketing of financial products) that no longer need to be filtered through parental control. Although many of these secular changes are particularly pronounced for adolescents living in high-income countries, the environment in low- and middle-income countries is rapidly changing to reproduce these phenomena, often in contexts in which there is less education and regulation to buffer their potentially deleterious effects (for example, the rapid changes in point of sale advertising for alcohol, tobacco and unhealthy snack foods[34]).

Given the importance of successful adolescent development for adult attainment of social, educational and financial success, achieving these goals will require policy initiatives and investments aimed at adolescents. A recent study has found evidence for high returns from such investments[7]. However, there is marked variation between countries, with returns expected to be greatest in low-income countries. A range of interventions, including policies targeting physical, mental and sexual health, road traffic injuries, child marriage, and increasing the extent and quality of secondary schooling, are all predicted to be associated with especially high benefit to cost ratios[7]. Accordingly, there have been many calls for comprehensive and increasing investments in adolescent health and wellbeing in both national and international policy[15,35,36].

We support these proposals to invest in the lives of adolescents, and moreover assert that in order to have the most positive impact, such investments must be guided by science. To ensure that policies are not too diffuse, they should also be carefully targeted to the epochs in which these policies may have the greatest impact.

Informing strategic investment in adolescence

Maturational periods of rapid growth and change can create opportunities for pivotal influences on developmental trajectories. This principle can be illustrated by rates in physical growth (see Fig. 1). For example, it is easy to anticipate the probable impact of experiencing a severe famine during a period of accelerated growth. Given the extra nutritional and metabolic demands of rapid growth, it makes sense that being an infant or being in the middle of a pubertal growth spurt during a period of malnutrition could have particularly deleterious long-term effects. Indeed, analyses of data from the famine during the Nigerian Civil war[37] and from the famine in Cambodia (1975–1979)[38] show that experiencing famine as an adolescent resulted in more pronounced stunting of adult height than does exposure to famine during other life stages. Another study has shown that being in early adolescence during historical periods of adverse experiences, such as war or famine, is particularly associated with a shortened life expectancy[39].

However, periods of rapid growth and maturational change can also create opportunities. For example, establishing healthy bone density during the adolescent growth spurt (by getting enough calcium, vitamin D and exercise) can be protective against developing osteoporosis in later life[40]. Similarly, adolescence creates an opportunity to normalize the risk trajectory of developing polycystic ovarian syndrome during adulthood among females with very low birth weight, through insulin sensitization (via metformin treatment) when administered during pubertal maturation[41].

These examples highlight a general principle about modifiable inflection points along developmental trajectories. We believe that there are parallels in adolescent development in cognitive, affective, social and motivational domains, which could provide opportunities to modify developmental trajectories in behavioural, educational and mental health outcomes. Peri-pubertal changes in learning and motivation and in the neural systems that underlie these changes may provide these opportunities.

Brain development and changes in learning

Both human and animal models have provided evidence for distinctive neurodevelopmental changes during adolescence (see Box 1). Most notably, the grey matter in higher-order brain regions becomes thinner and prunes synapses during adolescence[42–44]. However, some specific brain connections are formed primarily in early to mid-adolescence, which may potentiate new forms of motivation and and/or new learning. For example, dopaminergic[45] and amygdala neurons[46,47] increasingly innervate the frontal cortices, and the frontal cortex shows striking changes in top-down innervation of the amygdala[48]. There is a strengthening of functional connectivity between cortical regions implicated in cognitive control and the basal ganglia (relative to more affective regions), altering the ratio of cognitive versus affective inputs in a region that serves as an mediator of value-based decision-making[49]. Finally, gains in special inhibitory synapses may enable bursts of neuroplasticity[10,50,51], which may function to transiently enhance learning from experience in a sensitive period (Box 1).

Mid- to late adolescence is also a period of more global refinement and stabilization for the brain. Although many of these processes start much earlier in development, mid- to late adolescence is particularly associated with slowing in the rate of thinning in the grey matter regions implicated in higher cognition[52], decline in the outgrowth of new cortical synapses[47,53], and decreases in the total density of synapses on pyramidal neurons[43,44]. Also, myelination increases[54,55] and perineuronal nets accumulate in the extracellular space around some neurons[50,56]. Together, these processes are thought to bring greater stabilization and efficiency that may enhance function but could potentially limit further change[50,56–58].

Overall, these diverse changes across brain systems in adolescence result in a series of shifts in how the brain attends to, integrates and retains information[6,59]. This means that changes in the brain during adolescence not only shape behaviour, but also learning—in ways that could have a lifelong impact. Across development, our capacity for learning does not just gradually (and monolithically) increase, but rather learning and motivation change in nonlinear ways, and can show

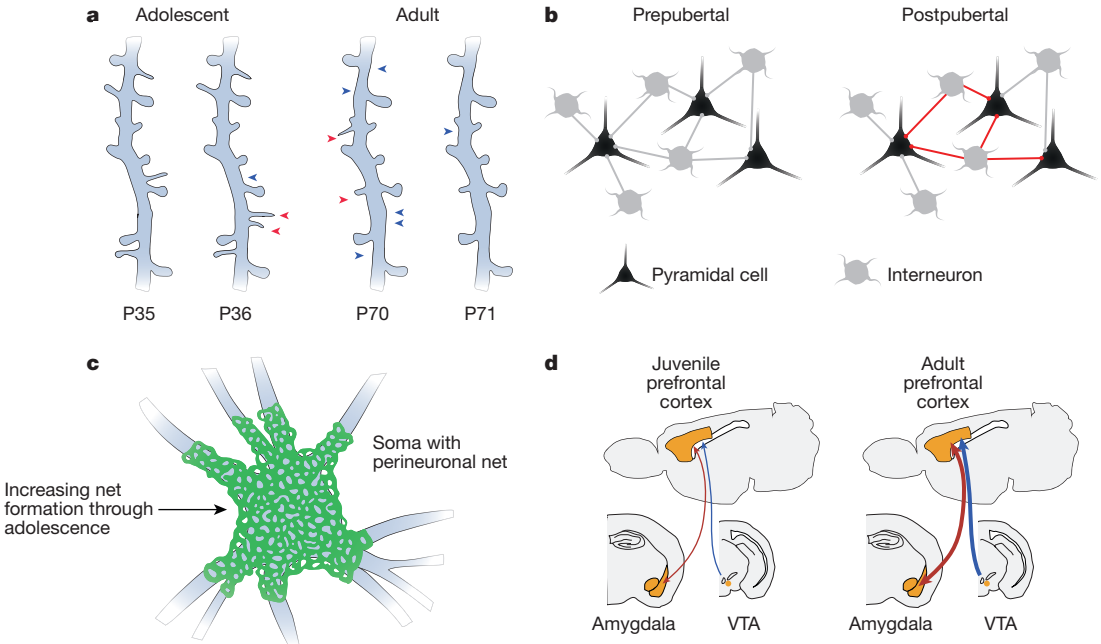
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

BOX 1

Changes in the adolescent brain at the level of circuits, cells and synapses.

The mammalian brain is remodelled during adolescence, particularly in associative cortical regions. Studies of human brains and animal models have shown that pruning of cortical synapses, as well as gaining of select new synapses in the neocortex, occurs during adolescence. These changes may reflect loss of juvenile forms of plasticity as well as potentially adolescent-specific sensitive periods. Whereas some of these changes have been shown to be puberty-dependent, others are thought to be puberty-independent. Shifts in pubertal timing could potentially alter the sequence and/or phase overlap of different maturational events.

In **a**, a schematic of a dendritic segment is shown at four time points, illustrating both the pruning and stabilization of spines (protrusions where synapses can form) that occur during adolescence. The first two time points are 24 h apart in early adolescence, whereas the second two are 24 h apart in early adulthood. When comparing adolescent time points to the adult time points, a decrease in spine density can be observed (lost spines are indicated by blue arrowheads). In addition, at the younger age, it has been shown that a greater fraction of the spines are lost or gained (red arrowheads) over 24 h, compared to the adult brain. These transient spines are thought to reflect the dendrite 'sampling' of the local space for potential new synaptic connections. As spine gains and losses become more infrequent, opportunities for rewiring and neuroplasticity are thought to lessen. Thus, during adolescence when dendritic spines become less dense and more stable, experience-dependent learning may be solidified, connectivity may become more efficient, and plasticity may become more limited. It remains debated whether spine changes are related to gonadal hormone effects at puberty[42,116]. In **b**, the schematic shows the increased inhibitory neurotransmission on to frontal cortical pyramidal neurons that occurs after puberty (indicated by red lines). This increase in a special form of neurotransmission has been shown to be organized by ovarian hormones at puberty in female mice[51]. The image in **b** is based on data from refs 10, 51. Changes in inhibition are thought to regulate the balance of excitation and inhibition and could initiate a transient sensitive period for plasticity[50]. A schematic of a perineuronal net surrounding the soma of a neuron is shown in **c**. Perineuronal nets increasingly form during adolescence in the neocortex[58] and amygdala[56], particularly around inhibitory neurons. These changes may limit plasticity and potentially sensitive periods across the brain[56,58] and in multiple species[57]. The schematic in **d** shows cross-sections of the rodent brain that highlight new long-range connections, which continue to form during adolescence. This is notable, because pruning of connectivity generally predominates in this phase (as in **a**). The arrows indicate increasing innervation of the prefrontal cortex by amygdala inputs[46,47] (red) and dopaminergic inputs from the ventral tegmental area (VTA)[45,69] (blue) during the adolescent transition. The schematic also shows that top-down prefrontal cortex innervation of the amygdala (red) undergoes marked growth during adolescence[48]. Human imaging studies have shown that similar prefrontal–amygdala remodelling may be occurring in human brains[117]. It has been shown that the increase in dopaminergic innervation of the prefrontal cortex, does not show sex-specific changes at pubertal milestones[118] but the role of amygdala connectivity during puberty has not been tested, to our knowledge.



qualitative as well as quantitative shifts. If we focus on highly specific forms of learning using carefully designed tasks, we can observe that some forms of learning are 'better' (meaning more accurate or more efficient) for adolescents[60–63], whereas other forms are 'better' for adults, or children[64–66].

There is considerable complexity in the science of learning during adolescence. Nevertheless, during early to mid-adolescence, there appears to be evidence of distinctive enhancements in: (1) exploration[64], sensation seeking and sensitivity to novelty[61,67–69]; (2) learning from situations in which information needs to be integrated over time from multiple experiences with imperfect (probabilistic) feedback[60,62,68] (a process that may be especially important for acquiring social, emotional and cultural competence, which cannot be easily mastered by memorizing rules or simply reasoning one's way through complex, rapidly changing social situations);(3) learning from negative feedback or about negative associations[62,68]; and (4) learning from social information obtained from peer interactions (both positive and negative)[4,70].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

In particular, processing of social information and feedback may undergo distinctive developmental changes in adolescence. Reviews have previously described adolescence as a period of 'social reorientation'[4,9,12,71], which includes increases in sensitivity to social evaluation and the importance of social status and popularity[18]. The social reorientation of adolescents is thought to be associated with changes in neural systems that are involved in the development of social cognition[72]—especially affective responses to social acceptance and social rejection[73,74]—and neural changes that are thought to enhance the motivational importance of engaging with peers[75], which may motivate behaviour and serve to enhance learning. (While acknowledging this increasing importance of peer relationships, the quality of family relationships continues to have an important influence on adolescent development and health, and provides a compelling example of bidirectional influences between social and biological development. See Box 2.)

The role of the late maturation of the prefrontal cortex in these shifts in learning is probably multifaceted, supporting not only gains in function but also weakening contributions from other systems[49,60]. The role of pubertal hormones in adolescent brain and behavioural changes is also not yet fully characterized[10], but it is likely that puberty-dependent and puberty-independent components and processes are intermingled[69]. Gonadal hormones have long been recognized to have a role in subcortical development[5], but there is also evidence that pubertal processes play a critical part in neocortical maturation[69,76–78] and shifts in learning[51].

Animal models and sensitive periods for learning

It may be informative to consider the social learning that occurs during adolescence as the product of a potential sensitive period. Over time, evolutionary influences have shaped each species' development in order to regulate the timing of different forms of learning. Sensitive periods for learning are timed to when important information is available and most useful to the adaptive development of the organism[79–81]. Humans and laboratory animals have been found to share similar early sensitive periods in sensory regions, such as in binocular visual regions for which there is a critical balance in the input from two eyes[50]. Humans also have multiple early sensitive periods for the acquisition of language[79,82]. There may be additional undiscovered sensitive periods for human cognitive, affective, and, importantly, social development that could be leveraged to enhance the impact of interventions. We can use sensitive periods for learning and social behaviour in other species as a guide. Here we focus on the biology of the sensitive periods that may be associated with puberty onset.

Human social learning during adolescence may share some parallels with sensitive periods for the acquisition of song learning in song birds. The timing of song learning in songbirds differs between species and involves imprinting on a song model (a 'tutor') through social interactions[79]. Some birds learn their song from their father and go through puberty in the first 90 days of life[79]. Others time their learning (and gonadal hormone and receptor fluctuations) to allow song plasticity during different periods[79,83,84], for example, when they arrive in a new territory and should imitate the songs of their new neighbours[79,85]. Gonadal hormones are thought to regulate the neurobiology of song learning by altering both motor flexibility and the salience of social cues. Changes in salience are thought to ensure appropriate selection of the song to be copied (that is, the correct species, time and context)[86–89] and selection of an appropriate mate[86–91].

In mammals, frontal neocortical areas, which have a critical role in higher cognition, self-regulation and social behaviours, develop late—from the time of pubertal milestones and into early adulthood[10,45–47,77]. In laboratory mice, the onset of puberty has been found to rapidly alter frontal cortex neurobiology, changing inhibitory neurotransmission mechanisms that have previously been identified as key regulators of sensitive periods in the neocortex[10,51] (see Box 1). Changes in frontal cortex inhibitory neurotransmission can also be connected to changes in learning[10,92], suggesting a causal link between puberty and adolescent shifts in learning.

In rats, the adolescent period is characterized by social learning, such as rough-and-tumble play and learning about sexual encounters. Interestingly, male rats need exposure to testosterone during adolescence in order to display rapid learning from sexual experience, and do not recover this learning if testosterone is replaced in adulthood[5,93]. Thus, in rodents, as in birdsong, we can see a repeated pattern, in which developmental signals (specifically, here, involving a change in gonadal hormone signalling) may prime the brain's ability to learn during an adolescent sensitive period when (1) the information is available in the environment and (2) developmentally appropriate. When the temporal intersection of these processes is prevented, which can occur when a developing adolescent is deprived of critical learning opportunities, the trajectory towards developing the capacities that are required to take up independent adult roles can be derailed or disrupted[5,79,93].

BOX 2

The role of parenting in adolescent social learning, health and brain development

Despite the strong cultural belief that the importance of parental contributions wanes compared to the input of peers during adolescence, empirical research has demonstrated that parenting is a strong determinant of adolescent health and well-being during this period of life, often more so than peer processes[119,120]. Moreover, recent empirical research has suggested that parenting processes during adolescence provide a compelling example of bidirectional relationships between environmental learning and/or experiences and biological development, potentially generating cascading effects of either vulnerability or resilience. For example, individual differences in adolescent brain development are associated with adolescents' behaviour during parent–child interactions[121], suggesting that neurobiology may influence interpersonal behaviour in family contexts. However, early adolescent parent–child relationships also prospectively predict patterns of future adolescent brain development[122], and brain function has been shown to serve a mediating link between family relationships and future psychopathology[123]. Finally, the quality of the parent–child relationship during early adolescence has been shown to buffer some of the potentially deleterious effects of low socio-economic status earlier in life on adolescent brain development[124]. Family processes during early adolescence are therefore potentially influenced by, and influence, both neurobiological development and functional outcomes. Of particular note is the potential role of family relationships as modifiable aspects of developmental processes during a maturational window of dynamic change. Greater precision in understanding the specific maturational changes (such as pubertal maturational changes in social information processing, and motivational salience) could provide insights into developmental timing, targets, and processes that have critical roles in these mutually interacting developmental processes in ways that inform leverage points for adolescent-focused public policy and clinical interventions. For example, although recent meta-analytic reviews have supported the role of parental intervention in the prevention of anxiety and depression[125], effect sizes are small, and few of these approaches have been informed by an explicitly developmental science perspective on the timing and targets of these interventions.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Developmental timing, investments and interventions

If comparable sensitive period processes occur in association with pubertal maturation in humans, this could inform developmental timing and targets for interventions. For example, if pubertal hormones underpin distinct changes in social learning, leveraging these insights by designing interventions that target this window of opportunity could have large positive effects on developmental trajectories.

The sharp rise in testosterone at the onset of puberty in both boys and girls is associated with shifts in social and affective information processing[67,94–96] and increased prioritization of social-status feedback[94]. A recent review of decades of research into the behavioural effects of testosterone in animals and humans has concluded that testosterone increases 'motivation to gain social status'[97]. Consistent with this conclusion, testosterone administration can increase status-enhancing behaviours in both prosocial and antisocial ways[98]. In humans, the process of earning prestige may shape behaviour in a manner that is separate from gaining status through aggressive dominance[99]. The specific types of learning—and criteria for earning prestige—relevant to social success are highly variable. Importantly, these are not only, or even primarily, determined by our biology.

A unique feature of our species, and potentially a distinct leverage point for intervention, is the role of human culture in shaping social learning relevant to prestige and admiration. Cultural values for prestige vary widely. For example, a Tibetan Buddhist community that confers the highest prestige on kindness and compassion will create different social learning opportunities for admiration-seeking adolescents than a warrior society or a highly materialistic culture. Developmental science cannot yet fully explain how and when these values are learned and maintained at the neural level. However, there appears to be promising evidence supporting these as potential leverage points for positive change during adolescence.

When designing interventions, it is important to note that adolescents are not simply passive learners conforming to adult values. Complaints about the rebellious contempt of adolescents towards adult authority go back to the time of Socrates. Although adolescents are strongly influenced by family and cultural values, they are also naturally seeking greater autonomy and independence. Dutifully acquiring knowledge and skills from accomplished adults represents one path for gaining status and prestige. An alternative pathway is to discover novel approaches or simply an innovative twist or refinement that leads to success. These tendencies are evident today as adolescents worldwide are early adopters of expertise with new technology (as well as frequent contributions to the leading edge of the latest trends in music, fashion and innovative use of language).

From this perspective, adolescence can be considered an inflection point for pivotal changes in human development in a very broad sense. Adolescent risk taking, together with status seeking (both of which are often maligned), may actually have contributed to unique group level capacities to exploit and preserve serendipitous innovations throughout human history. An extended (adolescent) developmental period of exploratory social learning—when an individual is also highly sensitive to social evaluation and the motivational importance of earned prestige through valued contribution—could have had a critical role in humans' remarkable capacity for rapid expansion, adoption and refinement of innovation through human culture. We speculate that interventions, which support and provide social scaffolding for healthy pro-social versions of innovation and success during this sensitive window of social learning and adolescent identity development, could potentially have large positive impacts on health and education.

Policy and intervention

The developmental science of adolescence is advancing rapidly, and we have only reviewed a small proportion of the field. We have selected work that illustrates the growing understanding of adolescence as a period of critical investment opportunity because of the specific types of learning that are potentiated during this period of development. In addition to scientific advances in understanding, the enormous potential for growth and positive development, we have also focused on areas of scientific progress that are providing insights into opportunities for specific policies and interventions. The emerging evidence points towards investments that place a strong emphasis on creating mastery learning experiences that maximize social learning and enhance status and autonomy at a key time in an individual's development of social identity and competence.

Translating our understanding of developmental processes to target specific outcomes will, of course, require careful attention to contextual norms around sex, gender and culture. As an example, let us further explore how an enhanced understanding of adolescent sensitivity to social evaluation and the motivational importance of earning prestige might create a window of opportunity for positive social learning experience to address gender inequalities. Consider, for example, the harmful social learning experiences for girls coming into adolescence in a social context that admires bold, courageous behaviour in boys, but disapproves of these qualities in girls. By contrast, girls who attend schools or clubs in which prestige and admiration signals are comparable across genders could have very different social learning experiences[100]. Given that puberty in girls typically starts between 10 and 12 years of age, this may be a particularly important time to promote gender equality by providing opportunities for girls to discover, through experiential learning, that they can excel and gain admiration and prestige for their bold successes (at an age when they are also typically bigger, stronger and more socially competent than boys of the same age). A formative learning experience during this key developmental window is likely to also positively influence the well-established gender differences in rates of mental disorders, which typically emerge during adolescence and which are also likely to be related to differences in the social experiences of girls and boys during these formative periods in early adolescence[101,102].

The perspective we present on prestige learning also provides a novel approach to risk taking tendencies in adolescence—often considered the *sine qua non* of unhealthy adolescent behavioural tendencies. Research has shown that adolescents often find risk-taking behaviours reinforcing, and moreover, that such behaviours can enhance their social reputation[103,104]. Indeed, courage is a quality that is admired almost uniformly across human cultural contexts, and showing courage in ways that are aligned to cultural values is a powerful route to earning prestige. As promoted through positive youth development frameworks, having pro-social opportunities to demonstrate courage—through sports, drama, civic engagement or supporting social justice—are likely to have enhanced positive effects during this developmental period[105]. Such experiences may not only prevent antisocial and self-injurious paths, but may also promote healthy trajectories and identity development. Moreover, the tendency to take risks—especially if it creates gains in status or prestige—now affects adolescents in new ways, as they can post videos of daring acts or innovative creation online and get thousands of 'likes' from all over the world. Social media amplify the social reinforcement associated with risk-taking behaviour.

Regarding more traditional approaches to prevention and early intervention for health problems in adolescence, a previous review[106] has focused on the relative failure of many traditional behavioural interventions during adolescence, because well-intentioned efforts to instruct adolescents to make positive changes to improve their health often indirectly (or directly) imply that adolescents require adult expertise and are unable to make the right choices on their own. Interventions that honour adolescents' sensitivity to status and respect appear to more effectively capture adolescent attention and motivation and can result in improved behavioural outcomes[106]. These same developmental science-based insights have implications for education policy—especially for students aged 10–14. The importance of student-driven learning (autonomy), collaborative learning (social engagement in learning) and school and classroom climates that honour adolescent sensitivities to status, respect and purpose, are likely to have powerful positive effects on learning—particularly in this formative period[106].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Table 1 | Examples of matching intervention strategies to developmental changes during adolescence

Developmental changes	Intervention strategies
Period of rapid physical growth and increased metabolic demands	•Mitigate risk of famine and malnutrition[38] •Establish a healthy bone density through optimal calcium and vitamin D intake and exercise[40]
Increased tendency to explore, seek novelty and excitement	•Create opportunities for positive risk taking that results in healthy, positive, productive, high arousal learning[103,104] •Provide context for self-directed, discovery/exploratory learning
Pubertal changes in sleep and circadian regulation	•Couple later school start times with prevention and intervention efforts to improve and regularize sleep[113] •Provide sleep interventions to youth at increased risk for mental-health problems[114]
Motivation for status, prestige and respect	•Design health promotion and behavioural interventions in ways that honour adolescents' sensitivity to autonomy, respect and prestige[106] •Train adult teachers/facilitators to treat adolescents with respect and appropriately enhance autonomy[115]
Motivation for social learning	•Leverage social relationships to reinforce positive behaviour[94,98] •Support adolescents in processing social information and understanding social experiences •Structure learning experiences to include, rather than ignore, social context[70,72] •Enhance collaborative social learning in education settings[19]
Identity development and heartfelt goals	•Introduce positive role models and mentors who exemplify feeling-based values and inspired goals[23] •Create mastery curve learning opportunities that gradually become more challenging

Applying a developmental science perspective to strategic investment and intervention requires consideration of a number of principles. Firstly, developmental epidemiology[107] can provide information on which opportunities and/or problems arise during different life stages, and can, therefore, provide suggestions regarding appropriate, developmentally informed prevention and intervention targets. Secondly, there should be an understanding of an underlying developmental process that results in greater sensitivity to interventions at this stage of life[106,108]. Finally, the process must be modifiable though educational, clinical, public health, policy or other interventions and have protective effects. In this sense, the developmental perspective can contribute not only to precision medicine[109] and precision public health[110], but also to precision public policy, by identifying ways to target investment and intervention for maximum benefit. Bearing in mind these three principles, Table 1 presents examples of developmental processes that putatively show sensitive periods during adolescence, along with examples of developmentally informed investment and intervention strategies. These are, however, only examples of potential opportunities. Other potential targets could include substance misuse, mental health, parenting, sexual and romantic relationships, nutrition, physical activity and use of digital technologies, to name a few.

As developmental science progresses to provide a more mechanistic understanding of developmental processes, it will lead to greater precision in understanding developmental risk factors and identifying the timing, mechanistic targets and best contexts to improve adolescent trajectories. This area of research is challenging, as it requires the evaluation of the differential effects of similar learning interventions across multiple developmental phases. However, some promising examples of such studies have been published[111,112]. Ultimately, the full integration of adolescent developmental science, intervention science and public policy will require a comprehensive evidence base that thoroughly evaluates which kinds of interventions can be improved upon through optimal matching to specific developmental windows of opportunity.

Outlook

The developmental science of adolescence is providing new insights into windows of opportunity during which we can have especially strong positive impacts on trajectories of health, education, social and economic success across the lifespan. This emerging science points towards adolescence as a time of enhanced growth and a sensitive period for learning—one in which adolescents' sensitivity to belonging, feeling valued and respected and finding a way to make a valued contribution (that is, to earn prestige and admiration) is also linked to adolescents' search for meaning and larger purpose. This social and affective learning can shape the development of 'heartfelt' goals and priorities, such as those associated with experiences of inspiration, creativity and innovation. Given current global changes, and the specific challenges affecting adolescents, there is a compelling need to understand the potential for strategic investments in adolescents to unleash this potential. From a global perspective, strategic investment of very limited resources, and the potential acceleration of economic inequalities and amplification of vulnerabilities and opportunities by the information technology revolution, makes the integration of insights from developmental science into public policy even more compelling. The stakes of investing in adolescents are rapidly increasing—especially if we are going to create sustainable growth, address climate change and reduce social inequalities—all problems that are critical for the achievement of the United Nations sustainable development goals. Strategic and developmentally informed investments in adolescents could contribute to a positive impact on the adolescents themselves, their future lives as leaders in adult society, and the next generation to whom they will be parents[7,15].

Received 14 November 2017; accepted 23 January 2018.

1. Shonkoff, J. P. & Garner, A. S. The lifelong effects of early childhood adversity and toxic stress. *Pediatrics* **129,** e232–e246 (2012).
2. Black, M. M. & Hurley, K. M. Investment in early childhood development. *Lancet* **384,** 1244–1245 (2014).
3. Shonkoff, J. P., Radner, J. M. & Foote, N. Expanding the evidence base to drive more productive early childhood investment. *Lancet* **389,** 14–16 (2017).
4. Crone, E. A. & Dahl, R. E. Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nat. Rev. Neurosci.* **13,** 636–650 (2012).
 This review of neuroimaging data presents a model of pubertal changes in social and affective processing, and how these may support the greater flexibility in motivations and priorities needed to navigate the changing social contexts of adolescence.
5. Schulz, K. M. & Sisk, C. L. The organizing actions of adolescent gonadal steroid hormones on brain and behavioral development. *Neurosci. Biobehav. Rev.* **70,** 148–158 (2016).
 This study reviews the organizational effects of gonadal hormones on brain and behaviour, and presents a 'wedge-shaped' model of decreasing sensitivity to gonadal hormonal effects with age.
6. Lee, F. S. *et al.* Adolescent mental health—opportunity and obligation. *Science* **346,** 547–549 (2014).
7. Sheehan, P. *et al.* Building the foundations for sustainable development: a case for global investment in the capabilities of adolescents. *Lancet* **390,** 1792–1806 (2017).
 This study estimates returns for investments in adolescents in low-income, lower-middle income, and upper-middle income countries, and concludes that investments in health and education could generate high economic and social returns.
8. Patton, G. C. & Viner, R. Pubertal transitions in health. *Lancet* **369,** 1130–1139 (2007).
9. Nelson, E. E., Jarcho, J. M. & Guyer, A. E. Social re-orientation and brain development: An expanded and updated view. *Dev. Cogn. Neurosci.* **17,** 118–127 (2016).
10. Piekarski, D. J. *et al.* Does puberty mark a transition in sensitive periods for plasticity in the associative neocortex? *Brain Res.* **1654,** 123–144 (2017).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

11. De Lorme, K., Bell, M. R. & Sisk, C. L. The teenage brain: social reorientation and the adolescent brain—the role of gonadal hormones in the male Syrian hamster. *Curr. Dir. Psychol. Sci.* **22,** 128–133 (2013).
12. van den Bos, W. Neural mechanisms of social reorientation across adolescence. *J. Neurosci.* **33,** 13581–13582 (2013).
13. Crockett, L. J. & Crouter, A. C. *Pathways through Adolescence: Individual Development in Relation to Social Contexts* (Lawrence Erlbaum, 1994).
14. Cohen, A. O. *et al.* When is an adolescent an adult? Assessing cognitive control in emotional and nonemotional contexts. *Psychol. Sci.* **27,** 549–562 (2016).
15. Patton, G. C. *et al.* Our future: a *Lancet* commission on adolescent health and wellbeing. *Lancet* **387,** 2423–2478 (2016).
This comprehensive report describes how unprecedented global forces are shaping the health and wellbeing of the largest generation of adolescents in human history.
16. Cohen, A. O., Bonnie, R. J., Taylor-Thompson, K. & Casey, B.J. When does a juvenile become an adult: implications for law and policy. *Temp. Law Rev.* **88,** 769–788 (2015).
17. Davey, C. G., Yücel, M. & Allen, N. B. The emergence of depression in adolescence: development of the prefrontal cortex and the representation of reward. *Neurosci. Biobehav. Rev.* **32,** 1–19 (2008).
18. Blakemore, S.-J. & Mills, K. L. Is adolescence a sensitive period for sociocultural processing? *Annu. Rev. Psychol*. **65,** 187–207 (2014).
This study describes how brain regions involved in social processing and behaviour undergo both structural changes and functional reorganization during the second decade of life, possibly reflecting a sensitive period for adapting to one's social environment.
19. Blakemore, S. J. & Choudhury, S. Development of the adolescent brain: implications for executive function and social cognition. *J. Child Psychol. Psychiatry* **47,** 296–312 (2006).
20. Dumontheil, I. Development of abstract thinking during childhood and adolescence: the role of rostrolateral prefrontal cortex. *Dev. Cogn. Neurosci*. **10,** 57–76 (2014).
21. de Water, E., Cillessen, A. H. & Scheres, A. Distinct age-related differences in temporal discounting and risk taking in adolescents and young adults. *Child Dev.* **85,** 1881–1897 (2014).
22. Gestsdottir, S., Bowers, E., von Eye, A., Napolitano, C. M. & Lerner, R. M. Intentional self regulation in middle adolescence: the emerging role of loss-based selection in positive youth development. *J. Youth Adolesc.* **39,** 764–782 (2010).
23. Schwartz, S. J. & Petrova, M. Fostering healthy identity development in adolescence. *Nat. Hum. Behav*. https://doi.org/10.1038/s41562-017-0283-2 (2018).
24. Suleiman, A. B., Galván, A., Harden, K. P. & Dahl, R. E. Becoming a sexual being: the 'elephant in the room'of adolescent brain development. *Dev. Cogn. Neurosci*. **25,** 209–220 (2017).
This review explores the role of puberty in the social, emotional and cognitive maturation processes necessary for reproductive success, and the role of sex and romance as important developmental dimensions of health and well-being in adolescence.
25. Dahl, R. E. Adolescent brain development: a period of vulnerabilities and opportunities. Keynote address. Ann. *NY Acad. Sci.* **1021,** 1–22 (2004).
26. United Nations. *World Population Prospects, the 2012 Revision.* https://esa.un.org/unpd/wpp/ (UNDESA, 2013).
27. Bell, V., Bishop, D. V. & Przybylski, A. K. The debate over digital technology and young people. *Br. Med. J.* **351,** h3064 (2015).
28. Madden, M., Lenhart, A., Duggan, M., Cortesi, S. & Gasser, U. *Teens and Technology 2013* http://www.pewinternet.org/2013/03/13/teens-and-technology-2013/ (2016).
29. Spies Shapiro, L. A. & Margolin, G. Growing up wired: social networking sites and adolescent psychosocial development. *Clin. Child Fam. Psychol. Rev.* **17,** 1–18 (2014).
30. Smetana, J. G., Metzger, A., Gettman, D. C. & Campione-Barr, N. Disclosure and secrecy in adolescent–parent relationships. *Child Dev.* **77,** 201–217 (2006).
31. Notten, N., Peter, J., Kraaykamp, G. & Valkenburg, P. M. Research note: digital divide across borders—a cross-national study of adolescents' use of digital technologies. *Eur. Sociol. Rev.* **25,** 551–560 (2009).
32. Worthman, C. M. & Trang, K. Dynamics of body time, social time and life history at adolescence. *Nature* https://doi.org/10.1038/nature25750 (2018).
33. Palan, K. M., Gentina, E. & Muratore, I. Adolescent consumption autonomy: a cross-cultural examination. *J. Bus. Res.* **63,** 1342–1348 (2010).
34. Stuckler, D., McKee, M., Ebrahim, S. & Basu, S. Manufacturing epidemics: the role of global producers in increased consumption of unhealthy commodities including processed foods, alcohol, and tobacco. *PLoS Med*. **9,** e1001235 (2012).
35. McCarthy, K. B. M. & Hall, K. *Investing when it Counts: Reviewing the Evidence and Charting a Course of Research and Action for Very Young Adolescents* (Population Council, 2016).
This report reviews the literature and presents a compelling case that investing in very young adolescents (ages 10–14) is important, as this is a pivotal time in the life course.
36. World Health Organization. Global Accelerated Action for the Health of Adolescents (AA-HA!): guidance to support country implementation. http://www.who.int/maternal_child_adolescent/topics/adolescence/framework-accelerated-action/en/ (WHO, 2017).
37. Akresh, R., Bhalotra, S., Leone, M. & Osili, U. O. War and stature: growing up during the Nigerian Civil War. *Am. Econ. Rev.* **102,** 273–277 (2012).
38. Holmqvist, G. & Pereira. A. Famines and stunting: Are adolescents the hardest hit? https://blogs.unicef.org/evidence-for-action/famines-and-stunting-are-adolescents-the-hardest-hit/ (Unicef, 2017).
39. Falconi, A., Gemmill, A., Dahl, R. E. & Catalano, R. Adolescent experience predicts longevity: evidence from historical epidemiology. *J. Dev. Orig. Health Dis*. **5,** 171–177 (2014).
40. Bonjour, J. P., Theintz, G., Buchs, B., Slosman, D. & Rizzoli, R. Critical years and stages of puberty for spinal and femoral bone mass accumulation during adolescence. *J. Clin. Endocrinol. Metab.* **73,** 555–563 (1991).
41. Mericq, V. *et al.* Long-term metabolic risk among children born premature or small for gestational age. *Nat. Rev. Endocrinol.* **13,** 50–62 (2017).
42. Drzewiecki, C. M., Willing, J. & Juraska, J. M. Synaptic number changes in the medial prefrontal cortex across adolescence in male and female rats: a role for pubertal onset. *Synapse* **70,** 361–368 (2016).
43. Petanjek, Z. *et al.* Extraordinary neoteny of synaptic spines in the human prefrontal cortex. *Proc. Natl Acad. Sci. USA* **108,** 13281–13286 (2011).
44. Rakic, P., Bourgeois, J. P. & Goldman-Rakic, P. S. Synaptic development of the cerebral cortex: implications for learning, memory, and mental illness. *Prog. Brain Res.* **102,** 227–243 (1994).
45. Benes, F. M., Vincent, S. L., Molloy, R. & Khan, Y. Increased interaction of dopamine-immunoreactive varicosities with GABA neurons of rat medial prefrontal cortex occurs during the postweanling period. *Synapse* **23,** 237–245 (1996).
46. Cunningham, M. G., Bhattacharyya, S. & Benes, F. M. Amygdalo-cortical sprouting continues into early adulthood: implications for the development of normal and abnormal function during adolescence. *J. Comp. Neurol.* **453,** 116–130 (2002).
47. Johnson, C. M. *et al.* Long-range orbitofrontal and amygdala axons show divergent patterns of maturation in the frontal cortex across adolescence. *Dev. Cogn. Neurosci*. **18,** 113–120 (2016).
48. Arruda-Carvalho, M., Wu, W. C., Cummings, K. A. & Clem, R. L. Optogenetic examination of prefrontal–amygdala synaptic development. *J. Neurosci.* **37,** 2976–2985 (2017).
49. Larsen, B., Verstynen, T. D., Yeh, F. C. & Luna, B. Developmental changes in the integration of affective and cognitive corticostriatal pathways are associated with reward-driven behavior. *Cereb. Cortex* https://doi.org/10.1093/cercor/bhx162 (2017).
50. Hensch, T. K. Critical period plasticity in local cortical circuits. *Nat. Rev. Neurosci*. **6,** 877–888 (2005).
51. Piekarski, D. J., Boivin, J. R. & Wilbrecht, L. Ovarian hormones organize the maturation of inhibitory neurotransmission in the frontal cortex at puberty onset in female mice. *Curr. Biol*. **27,** 1735–1745 (2017).
This study provides evidence for organizational effects of ovarian hormones on the maturation of inhibitory neurotransmission in the cingulate cortex in mice, and provides experimental evidence for puberty-linked changes in learning.
52. Tamnes, C. K. *et al.* Development of the cerebral cortex across adolescence: a multisample study of inter-related longitudinal changes in cortical volume, surface area, and thickness. *J. Neurosci.* **37,** 3402–3412 (2017).
53. Zuo, Y., Lin, A., Chang, P. & Gan, W. B. Development of long-term dendritic spine stability in diverse regions of cerebral cortex. *Neuron* **46,** 181–189 (2005).
54. Baum, G. L. *et al.* Modular segregation of structural brain networks supports the development of executive function in youth. *Curr. Biol.* **27,** 1561–1572 (2017).
55. Paus, T. Growth of white matter in the adolescent brain: myelin or axon? *Brain Cogn.* **72,** 26–35 (2010).
56. Gogolla, N., Caroni, P., Lüthi, A. & Herry, C. Perineuronal nets protect fear memories from erasure. *Science* **325,** 1258–1261 (2009).
57. Balmer, T. S., Carels, V. M., Frisch, J. L. & Nick, T. A. Modulation of perineuronal nets and parvalbumin with developmental song learning. *J. Neurosci.* **29,** 12878–12885 (2009).
58. Lee, H. H. C. *et al.* Genetic Otx2 mis-localization delays critical period plasticity across brain regions. *Mol. Psychiatry* **22,** 680–688 (2017).
59. Guskjolen, A., Josselyn, S. A. & Frankland, P. W. Age-dependent changes in spatial memory retention and flexibility in mice. *Neurobiol. Learn. Mem.* **143,** 59–66 (2017).
60. Davidow, J. Y., Foerde, K., Galván, A. & Shohamy, D. An Upside to reward sensitivity: the hippocampus supports enhanced reinforcement learning in adolescence. *Neuron* **92,** 93–99 (2016).
61. Johnson, C. & Wilbrecht, L. Juvenile mice show greater flexibility in multiple choice reversal learning than adults. *Dev. Cogn. Neurosci.* **1,** 540–551 (2011).
62. van den Bos, W., Cohen, M. X., Kahnt, T. & Crone, E. A. Striatum–medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. *Cereb. Cortex* **22,** 1247–1255 (2012).
63. van der Schaaf, M. E., Warmerdam, E., Crone, E. A. & Cools, R. Distinct linear and non-linear trajectories of reward and punishment reversal learning during development: relevance for dopamine's role in adolescent decision making. *Dev. Cogn. Neurosci*. **1,** 578–590 (2011).
64. Gopnik, A. *et al.* Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proc. Natl Acad. Sci. USA* **114,** 7892–7899 (2017).
65. Lucas, C. G., Bridgers, S., Griffiths, T. L. & Gopnik, A. When children are better (or at least more open-minded) learners than adults: developmental differences in learning the forms of causal relationships. *Cognition* **131,** 284–299 (2014).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

66. Walker, C. M., Bridgers, S. & Gopnik, A. The early emergence and puzzling decline of relational reasoning: effects of knowledge and search on inferring abstract concepts. *Cognition* **156,** 30–40 (2016).
67. Braams, B. R., van Duijvenvoorde, A. C., Peper, J. S. & Crone, E. A. Longitudinal changes in adolescent risk-taking: a comprehensive study of neural responses to rewards, pubertal development, and risk-taking behavior. *J. Neurosci.* **35,** 7226–7238 (2015).
68. Hauser, T. U., Iannaccone, R., Walitza, S., Brandeis, D. & Brem, S. Cognitive flexibility in adolescence: neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. *Neuroimage* **104,** 347–354 (2015).
69. Walker, D. M. *et al.* Adolescence and reward: making sense of neural and behavioral changes amid the chaos. *J. Neurosci.* **37,** 10855–10866 (2017).
This study describes the role pubertal hormones have in the development of adolescent social and reward-related behaviours with a focus on sex differences, the medial prefrontal cortex, mesocorticolimbic dopamine and amygdala in rodents.
70. Pfeifer, J. H. *et al.* Longitudinal change in the neural bases of adolescent social self-evaluations: effects of age and pubertal development. *J. Neurosci.* **33,** 7415–7419 (2013).
71. Nelson, E. E., Leibenluft, E., McClure, E. B. & Pine, D. S. The social re-orientation of adolescence: a neuroscience perspective on the process and its relation to psychopathology. *Psychol. Med.* **35,** 163–174 (2005).
72. Goddings, A. L., Burnett Heyes, S., Bird, G., Viner, R. M. & Blakemore, S. J. The relationship between puberty and social emotion processing. *Dev. Sci.* **15,** 801–811 (2012).
73. Somerville, L. H. *et al.* The medial prefrontal cortex and the emergence of self-conscious emotion in adolescence. *Psychol. Sci.* **24,** 1554–1562 (2013).
74. Stroud, L. R. *et al.* Stress response and the adolescent transition: performance versus peer rejection stressors. *Dev. Psychopathol.* **21,** 47–68 (2009).
75. Rosen, M. L. *et al.* Salience network response to changes in emotional expressions of others is heightened during early adolescence: relevance for social functioning. *Dev. Sci.* https://doi.org/10.1111/desc.12571 (2017).
76. Herting, M. M. & Sowell, E. R. Puberty and structural brain development in humans. *Front. Neuroendocrinol.* **44,** 122–137 (2017).
77. Juraska, J. M. & Willing, J. Pubertal onset as a critical transition for neural development and cognition. *Brain Res.* 1654, 87–94 (2017).
78. Sisk, C. L. Hormone-dependent adolescent organization of socio-sexual behaviors in mammals. *Curr. Opin. Neurobiol.* **38,** 63–68 (2016).
79. Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22,** 567–631 (1999).
80. Fawcett, T. W. & Frankenhuis, W. E. Adaptive explanations for sensitive windows in development. *Front. Zool.* **12,** S3 (2015).
81. Insel, T. R. & Fernald, R. D. How the brain processes social information: searching for the social brain. *Annu. Rev. Neurosci.* **27,** 697–722 (2004).
82. Werker, J. F. & Hensch, T. K. Critical periods in speech perception: new directions. *Annu. Rev. Psychol.* **66,** 173–196 (2015).
83. Alvarez-Buylla, A. & Kirn, J. R. Birth, migration, incorporation, and death of vocal control neurons in adult songbirds. *J. Neurobiol.* **33,** 585–601 (1997).
84. Kirn, J., O'Loughlin, B., Kasparian, S. & Nottebohm, F. Cell death and neuronal recruitment in the high vocal center of adult male canaries are temporally related to changes in song. *Proc. Natl Acad. Sci. USA* **91,** 7844–7848 (1994).
85. Templeton, C. N. *et al.* Immediate and long-term effects of testosterone on song plasticity and learning in juvenile song sparrows. *Behav. Processes* **90,** 254–260 (2012).
86. Bottjer, S. W. & Johnson, F. Circuits, hormones, and learning: vocal behavior in songbirds. *J. Neurobiol.* **33,** 602–618 (1997).
87. Marler, P., Peters, S., Ball, G. F., Dufty, A. M. Jr & Wingfield, J. C. The role of sex steroids in the acquisition and production of birdsong. *Nature* **336,** 770–772 (1988).
88. Marler, P., Peters, S. & Wingfield, J. Correlations between song acquisition, song production, and plasma levels of testosterone and estradiol in sparrows. *J. Neurobiol.* **18,** 531–548 (1987).
89. Remage-Healey, L., Dong, S. M., Chao, A. & Schlinger, B. A. Sex-specific, rapid neuroestrogen fluctuations and neurophysiological actions in the songbird auditory forebrain. *J. Neurophysiol.* **107,** 1621–1631 (2012).
90. Matragrano, L. L., LeBlanc, M. M., Chitrapu, A., Blanton, Z. E. & Maney, D. L. Testosterone alters genomic responses to song and monoaminergic innervation of auditory areas in a seasonally breeding songbird. *Dev. Neurobiol.* **73,** 455–468 (2013).
91. Matragrano, L. L. *et al.* Estradiol-dependent modulation of serotonergic markers in auditory areas of a seasonally breeding songbird. *Behav. Neurosci.* **126,** 110–122 (2012).
92. Cho, K. K. *et al.* Gamma rhythms link prefrontal interneuron dysfunction with cognitive inflexibility in *Dlx5/6*$^{+/-}$ mice. *Neuron* **85,** 1332–1343 (2015).
93. De Lorme, K. C. & Sisk, C. L. The organizational effects of pubertal testosterone on sexual proficiency in adult male Syrian hamsters. *Physiol. Behav.* **165,** 273–277 (2016).
94. Cardoos, S. L. *et al.* Social status strategy in early adolescent girls: testosterone and value-based decision making. *Psychoneuroendocrinology* **81,** 14–21 (2017).
95. Spielberg, J. M., Olino, T. M., Forbes, E. E. & Dahl, R. E. Exciting fear in adolescence: does pubertal development alter threat processing? *Dev. Cogn. Neurosci.* **8,** 86–95 (2014).
96. Tyborowska, A., Volman, I., Smeekens, S., Toni, I. & Roelofs, K. Testosterone during puberty shifts emotional control from pulvinar to anterior prefrontal cortex. *J. Neurosci.* **36,** 6156–6164 (2016).
97. Bos, P. A., Panksepp, J., Bluthé, R. M. & van Honk, J. Acute effects of steroid hormones and neuropeptides on human social-emotional behavior: a review of single administration studies. *Front. Neuroendocrinol.* **33,** 17–35 (2012).
98. Dreher, J. C. *et al.* Testosterone causes both prosocial and antisocial status-enhancing behaviors in human males. *Proc. Natl Acad. Sci. USA* **113,** 11633–11638 (2016).
99. Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A. & Henrich, J. Two ways to the top: evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *J. Pers. Soc. Psychol.* **104,** 103–125 (2013).
100. Chandra-Mouli, V. *et al.* Implications of the global early adolescent study's formative research findings for action and for research. *J. Adolesc. Health* **61,** S5–S9 (2017).
101. Allen, N. B., Latham, M. D., Barrett, A., Sheeber, L. & Davis, B. in *Comprehensive Women's Mental Health* (eds Castle, D. & Abel, J.) 65–80 (Cambridge Univ. Press, 2016).
102. Avenevoli, S., Swendsen, J., He, J.-P., Burstein, M. & Merikangas, K. R. Major depression in the national comorbidity survey–adolescent supplement: prevalence, correlates, and treatment. *J. Am. Acad. Child Adolesc. Psychiatry* **54,** 37–44 (2015).
103. Maggs, J. L., Almeida, D. M. & Galambos, N. L. Risky business: the paradoxical meaning of problem behavior for young adolescents. *J. Early Adolesc.* **15,** 344–362 (1995).
104. Silbereisen, R. K. & Reitzle, M. in *Self-Regulatory Behaviour and Risk Taking: Causes and Consequences* (eds Lipsitt, L. P. & Mitnick, L. L.) 199–217 (Ablex, 1991).
105. Catalano, R. F., Hawkins, J. D., Berglund, M. L., Pollard, J. A. & Arthur, M. W. Prevention science and positive youth development: competitive or cooperative frameworks? *J. Adolesc. Health* **31,** 230–239 (2002).
106. Yeager, D. S., Dahl, R. E. & Dweck, C. S. Why interventions to influence adolescent behavior often fail but could succeed. *Perspect. Psychol. Sci.* **13,** 101–122 (2018).
This study provides a developmental perspective as to why traditional preventative school-based interventions work less well for adolescents, and reviews examples of promising approaches that take into account adolescents' enhanced desire to feel respected and be accorded status.
107. Costello, E. J. & Angold, A. in *Developmental Psychopathology* 3rd edn (ed. Cicchetti, D.) 1–35 (John Wiley & Sons, 2016).
108. Suleiman, A. B. & Dahl, R. E. Leveraging neuroscience to inform adolescent health: The need for an innovative transdisciplinary developmental science of adolescence. *J. Adolesc. Health* **60,** 240–248 (2017).
109. Collins, F. S. & Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* **372,** 793–795 (2015).
110. Khoury, M. J., Iademarco, M. F. & Riley, W. T. Precision public health for the era of precision medicine. *Am. J. Prev. Med.* **50,** 398–401 (2016).
111. Knoll, L. J. *et al.* A window of opportunity for cognitive training in adolescence. *Psychol. Sci.* **27,** 1620–1631 (2016).
112. Knoll, L. J., Leung, J. T., Foulkes, L. & Blakemore, S. J. Age-related differences in social influence on risk perception depend on the direction of influence. *J. Adolesc.* **60,** 53–63 (2017).
113. Minges, K. E. & Redeker, N. S. Delayed school start times and adolescent sleep: A systematic review of the experimental evidence. *Sleep Med. Rev.* **28,** 86–95 (2016).
114. Freeman, D. *et al.* The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. *Lancet Psychiatry* **4,** 749–758 (2017).
115. Okonofua, J. A., Paunesku, D. & Walton, G. M. Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proc. Natl Acad. Sci. USA* **113,** 5221–5226 (2016).
116. Anderson, S. A., Classey, J. D., Condé, F., Lund, J. S. & Lewis, D. A. Synchronous development of pyramidal neuron dendritic spines and parvalbumin-immunoreactive chandelier neuron axon terminals in layer III of monkey prefrontal cortex. *Neuroscience* **67,** 7–22 (1995).
117. Gabard-Durnam, L. J. *et al.* The development of human amygdala functional connectivity at rest from 4 to 23 years: a cross-sectional study. *Neuroimage* **95,** 193–207 (2014).
118. Willing, J., Cortes, L. R., Brodsky, J. M., Kim, T. & Juraska, J. M. Innervation of the medial prefrontal cortex by tyrosine hydroxylase immunoreactive fibers during adolescence in male and female rats. *Dev. Psychobiol.* **59,** 583–589 (2017).
119. Morris, A. S., Criss, M. M., Silk, J. S. & Houltberg, B. J. The impact of parenting on emotion regulation during childhood and adolescence. *Child Dev. Perspect.* **11,** 233–238 (2017).
120. Schwartz, O. S., Sheeber, L. B., Dudgeon, P. & Allen, N. B. Emotion socialization within the family environment and adolescent depression. *Clin. Psychol. Rev.* **32,** 447–453 (2012).
121. Whittle, S. *et al.* Prefrontal and amygdala volumes are related to adolescents' affective behaviors during parent–adolescent interactions. *Proc. Natl Acad. Sci. USA* **105,** 3652–3657 (2008).
122. Whittle, S. *et al.* Observed measures of negative parenting predict brain development during adolescence. *PLoS ONE* **11,** e0147774 (2016).
123. Callaghan, B. L. *et al.* Amygdala resting connectivity mediates association between maternal aggression and adolescent major depression: a 7-year longitudinal study. *J. Am. Acad. Child Adolesc. Psychiatry* **56,** 983–991 (2017).
124. Whittle, S. *et al.* Role of positive parenting in the association between neighborhood social disadvantage and brain development across adolescence. *JAMA Psychiatry* **74,** 824–832 (2017).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

125. Yap, M. B. *et al.* Parents in prevention: a meta-analysis of randomized controlled trials of parenting interventions to prevent internalizing problems in children from birth to age 18. *Clin. Psychol. Rev.* **50,** 138–158 (2016).
126. Richards, B. A. *et al.* Patterns across multiple memories are identified over time. *Nat. Neurosci.* **17,** 981–986 (2014).
127. Tompary, A. & Davachi, L. Consolidation Promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. *Neuron* **96,** 228–241 (2017).

Acknowledgements We thank the leadership team of the Center on the Developing Adolescent, including A. Galvan, A. Fuligni and J. Pfeifer, who have provided important intellectual contributions through many formative discussions over the past two years—in ways that were instrumental to an integrative understanding of the developmental science of adolescence as expressed in this paper.

Author Contributions R.E.D. and N.B.A. developed the outline of the paper. All authors drafted the manuscript, and provided critical revisions. All authors approved the final version of the manuscript for submission.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to N.B.A. (nallen3@uoregon.edu).

Reviewer Information *Nature* thanks B. J. Casey and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

PERSPECTIVE

doi:10.1038/nature25750

Dynamics of body time, social time and life history at adolescence

Carol M. Worthman[1] & Kathy Trang[1]

Recent opposing trends towards earlier physical maturation and later social maturation present a conundrum of apparent biological–social mismatch. Here we use life history analysis from evolutionary ecology to identify forces that drive these shifts. Together with findings in developmental science, our life history analysis indicates that adolescence is a distinctive period for biological embedding of culture. Ethnographic evidence shows that mass education is a novel feature of the globalizing cultural configurations of adolescence, which are driven by transformations in labour, livelihood and lifestyle. Evaluation of the life history trade-offs and sociocultural ecologies that are experienced by adolescents may offer a practical basis for enhancing their development.

For humans, the second decade of life is often described as a period of turmoil roiled by biological and sociocultural forces. This tension is captured in the linguistic distinction between puberty and adolescence, where puberty refers to the suite of physical changes comprising the transition from immaturity to maturity, juxtaposed against adolescence, which refers to the sociocultural construction of that transition. Dynamics behind these two forces are incompletely coupled, creating the potential for complex interactions generated by degrees of synchrony or dyssynchrony between them. Indeed, ongoing trends of globalization are transforming the timing of adolescence on both fronts: physical maturation has been accelerating even as sociocultural thresholds for achieving adulthood are rising[1,2].

These trends are well-recognized[3], but their origins and effects remain underexplored, hindering attempts to promote better outcomes across the board for adolescents. Recent conceptual and empirical advances in evolutionary ecology and in developmental science offer insight into factors that are driving these shifts in both biological timing and sociocultural schedules. Within evolutionary ecology, life history theory has identified trade-offs behind biological and sociocultural factors that influence timing of puberty and the configuration of adolescence[4,5]. In developmental science, ongoing identification of synergies between context and biology[6] points to adolescence as a period for biological embedding of culture, and highlights the potential impact of alterations in bio-contextual alignments[7]. These convergent insights come at a crucial time, when promotion of emerging adolescent capacities and health is vital for future economic and social prosperity, especially given global demographic ageing[2].

Here we describe human life history strategy, the determinants of physical and sociocultural timing of maturation, and the grounds for biological embedding of culture. Subsequently, we review evidence for acceleration in the timing of biological maturation, and evaluate it in the light of life history theory. Next, we delineate the changing cultural configurations of adolescence by contrasting patterns of historical ethnographic records with recent globalizing trends, and consider them in terms of life history strategy and the concept of embodied capital (see below). Lastly, we discuss the implications of these fresh insights for assumptions regarding globalizing formations of adolescence, for tracking life history trade-offs in development, and for considering social ecology as an approach to better meet the diverse developmental needs of adolescents that will help them to build viable futures. Recognition of adolescence as a critical window for the biological embedding of culture emphasizes the need for attention to interactions between pubertal maturation and ambient socioecological configurations that shape the emergent social–emotional and psychobehavioural capacities of adolescents.

Life history, puberty and adolescence

Life history theory aims to account for biodiversity, from ants to zebras, in terms of allocation of limited resources (time, energy) across the life course[8]. The challenge is to use these resources to carve out and inhabit a niche that sustains survival and reproduction[9]. A life history strategy represents an evolved, species-characteristic design for partitioning available resources among competing life demands (growth, reproduction or maintenance (metabolism and repair)) in order to optimize fitness within its ecological niche. Limited resources impose trade-offs among competing demands, and evolved life history strategies establish priorities and set trade-offs among these demands. Neuroendocrine processes negotiate trade-offs in resource allocation at the physiological and behavioural level, and scaffold the architecture of life history[10]. The schedule of trade-offs structures the timing of events, such as weaning, puberty, first birth or death, and outcomes, such as adult size, rate of reproduction or longevity.

Humans occupy a flexible niche characterized by skill-, technology- and cooperation-based extraction and sharing of high-quality resources, environmental modification and cultural complexes that support these behaviours and social arrangements[11]. The corresponding human life history strategy is slow: juvenile provisioning and other resource-intensive conditions enhance early survivorship and extend life expectancy, thus promoting protracted investment in development. Two distinctive human life history features are particularly relevant for adolescence, namely maturational delay and reliance on learning and cultural transmission. The extended juvenile period between infancy and adolescence accommodates demands for the formation of embodied capacities, including brain development, learning and socialization[12], and relies on deferring puberty. Extended childhood and delayed puberty are established by neuroendocrine mechanisms that repress the activity of the hypothalamic–pituitary–gonadal (HPG) reproductive axis a few months after birth[10].

Consequently, a coordinated set of mechanisms is required to derepress the HPG axis and permit puberty to proceed[13]. These mechanisms are sensitive to the quality of the environment (nutrition and pathogen burden), which determines trade-offs between advantages of continued growth versus gains from the onset of fertility[14]. Timing mechanisms provide adaptive flexibility to align maturation and onset of fertility with

[1]Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ambient resources and risk. Accordingly, growth rates and timing of puberty respond to resource availability and demands as well as mortality risk, accelerating under good conditions and slowing under poor ones[15]. As a result, human growth and age at puberty exhibit marked plasticity, or reaction norms, defined as the range of phenotypes that are expressed by a given genotype across disparate environmental conditions[4]. Thus, height may differ among countries by as much as 20 cm[16], and median age at menarche has been found to vary by more than six years from affluent urban to remote horticultural populations[1].

The human reliance on learning and cultural transmission underscores the importance of the large human brain and particularly the expanded human cortex, which is considered to be the basis of the cognitive and socio-emotional capacities that underlie human sociality, culture and adaptation[17]. The human brain undergoes a distinctive developmental pattern of extremely rapid growth in infancy, slower sustained growth in childhood and a final burst of growth and reorganization during the second decade of life that is initiated in puberty[18]. Infancy and childhood are recognized as important periods for learning and culture acquisition. However, accumulating imaging studies demonstrating important brain maturational changes in puberty and adolescence suggest that another learning period occurs during puberty and adolescence, when pubertal hormones are activated and cortical and subcortical structures of the brain are remodelled[19]. The pronounced structural and functional changes during this period follow sexually dimorphic trajectories[20,21]. Longitudinal studies of brain development through adolescence document not only anatomical changes[22,23], but also rapid advances in functional connectivity among brain regions[20] and emergence of individually distinctive connectivity patterns that also follow a sexually dimorphic course[24].

Discovery of these processes has stimulated intense investigation of how pubertal changes interact with life circumstances to shape social, emotional and behavioural development. Notably, they also create the conditions for biological embedding of cultural ecologies. Structural reorganization of the brain has been tied to functional changes in cognitive, affective and social processing in adolescence, although regional differences in cortical maturation rates may lead to temporal asynchronies in maturation of cognitive capacities. Crone and Dahl[25] have argued that dynamic asynchrony enables flexible engagement of cognitive control that is attuned to the motivational importance of contextual cues, including the presence of peers, perceived awards and threats, and task priority. Attunement to social context creates the basis for local socio-emotional development and learning. For instance, brain-imaging studies show that, relative to adults, adolescents have a stronger central neural response to threat and reward stimuli[26], to social exclusion[27,28] and to peer-observed task performance[29]. They also have an enhanced capacity to process evaluative feedback and modify behaviour accordingly[30]. Pubertal hormones (particularly testosterone) have been linked to these shifts in attention and motivation[31]. The heightened sensitivity to context and social evaluation, particularly by peers[32], supports the idea that adolescence is a sensitive period for specific forms of sociocultural learning and the acquisition of skillsets that are needed to transition successfully towards adult roles[33]. Heightened contextual sensitivity enhances responsiveness to cultural constructions of the second decade that strongly shape the contexts in which neurodevelopment and learning take place.

These dynamics modulate the continuous acquisition of cultural competence. Learning from others is the principal means for culture acquisition, however, some people are better sources than others[34,35]. How can learners decide with and from whom to learn? Infants swiftly assimilate huge chunks of the dominant culture (language, gestures and social behaviours) in everyday settings, but expanding mobility and social contacts with age raise options for selective updating. Puberty initiates an intense period of updating cultural information as maturational processes open up new domains and capacities (for example, romantic love, sexuality, moral logics and self-identity), triggering acute quests for relevant models and guides for direction and navigation as new skills are built and possible selves explored. Anthropological studies of cultural transmission have identified selective attention as a potent mediator of differential learning or biased transmission[36]. Learning may be frequency-dependent, biased towards normative (conformist bias) or rare (novelty bias) features: the former favours risk reduction and cultural stability, and the latter facilitates niche construction[9,37]. Learning may also be person-dependent, and sensitive to prestige markers of social status or success. People with perceived prestige or dominance, for instance, command more visual attention in experimental and naturalistic settings[36,38].

Adolescent attunement to social contexts, sensitivity to feedback and peer learning biases all suggest that maturational processes equip and motivate selective social learning and culture acquisition by adolescents. Reciprocally, cultural regulation of learning opportunities, such as assignment to gender- or class-based roles and settings, powerfully shapes development and acquisition of cultural competence[39]. These dynamics tailor the fit of adolescents to local social realities, such as performance of gender or occupational roles, yet may generate or perpetuate disparities and constrain life chances at adulthood.

Analyses of human life history have also identified embodied capital as a key developmental outcome[40]. Mere physical growth and maturation are insufficient for viability. Functional capacities are required, such as strength, immune function or resilience, as well as knowledge and skills that include linguistic, social and productive skills that are essential for subsistence and cultural competence[5]. Therefore, embodied or somatic capital comprises both physical and functional capabilities, the bases for meeting life challenges and opportunities. Humans invest heavily in embodied capital, both through prolonged growth and maturation, and through intensive concomitant investments in acquiring the functional capabilities that are necessary for productive, reproductive and social viability (knowledge, emotion regulation, executive function and social skills)[11]. Parents, and society as a whole, track the forms of embodied capital that will be needed for future success, and accordingly adjust their investments as well as the constructions of adolescence, including timing and criteria for adulthood[41]. Notably, cultural allocation of opportunities to build embodied capital both support access to life chances and may foster inequality, on the basis of gender, class, geographic or other grounds.

In summary, adolescence represents a pivotal point in life history when somatic investments in growth and maturation are completed and resources are reallocated to reproduction. Reproduction, however, entails not only the production of offspring, but also requires years of skilled investments in parenting, and the creation of contexts for the survival and nurturance of the young (ecological inheritance)[9]. Therefore, adolescence is a crucial period for generating embodied capital and not only in terms of reproductive or productive capacities, but also via pubertal processes that promote cultural embedding by preparing adolescents for context-sensitive formation or honing of socio-emotional skills and networks that are essential for their own and their future offsprings' survival.

Changes in the timing of physical maturation

Recent population changes match life history predictions, attest to wide reaction norms for growth rates and timing of puberty, and support their potential adaptive life history role in adjusting maturational timing to current environmental conditions[10]. In contrast to the obvious physical manifestations of puberty, such as growth or menarche, the underlying physiological processes are not easily analysed. Physiologically, puberty comprises a suite of changes that are orchestrated not only by the reproductive axis via gonadal steroids but also by adrenal, metabolic and other contributory factors[42]. Both casual and scientific assessments of pubertal timing and progress commonly rely on observable phenomena such as breast development, accelerated height gain or voice change[43–45]. Detectable signs of puberty can cue social and personal responses, whereas invisible albeit 'real' underlying conditions, such as gamete production, cannot. Menarche exemplifies this disjunction: an arresting and memorable event, it commonly acts as a marker of reproductive maturation, yet does not correspond to the first ovulation much less to established ovarian cyclicity[46]. Conversely, the absence of an equivalent

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

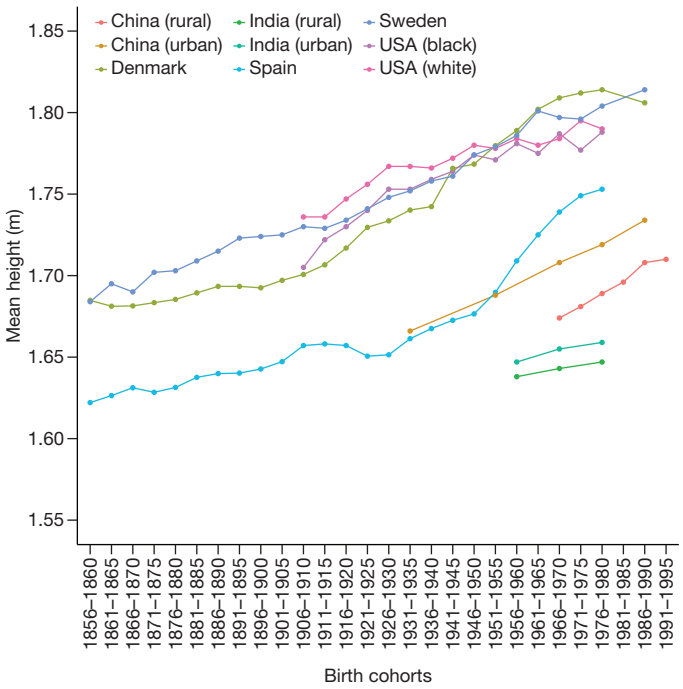


Figure 1 | Secular trends in male height. Drawn from data in ref. 97.

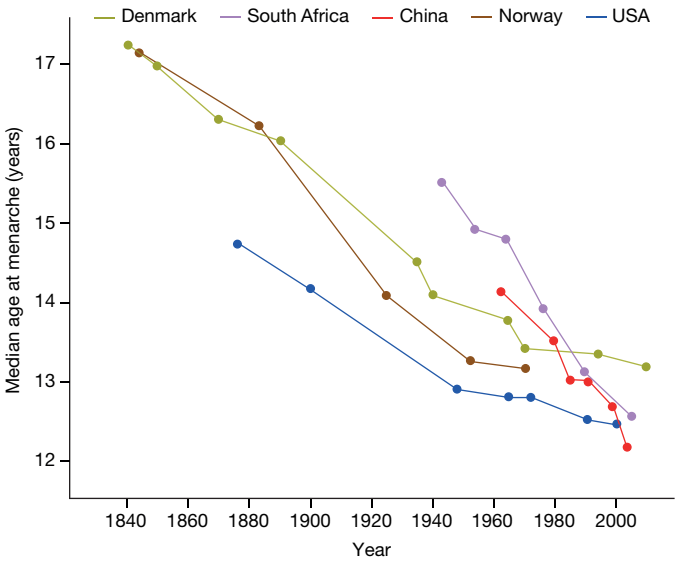


Figure 2 | Secular trends in age at menarche from first year of survey collection. Drawn from data in refs 50 and 98.

marker contributes to the more limited understanding of pubertal timing and variation among males.

Secular trends to increasing height and decreasing ages at menarche were first detected among European populations in the early nineteenth century[47] (Figs 1, 2). Declines were pronounced: median age at menarche decreased from over 17 to under 14 years among Norwegians from 1840 to 1940, for example. Subsequently, decreasing ages at menarche have been widely observed, but the timing and pace have varied among populations: decreases have proceeded more swiftly where they have occurred more recently[48,49], and appear to plateau around a median age of 12.5–13.5 years[50,51]. Continuing widespread acceleration in the onset of breast development (thelarche) contributes to perceptions that age at puberty continues to decrease, but appears to be dissociated from HPG activation and menarche, and may respond to environmental disruptors[52].

Tracking secular trends in male pubertal development has been complicated by the paucity of equivalent population studies and the use of indirect indicators, such as growth velocity, first seminal emission, voice break and risk-taking behaviour, that represent different aspects of puberty[44,45,53,54]. Studies that specifically examined pubic hair (pubarche) and testicular (gonadarche) development in boys have produced mixed results[55,56]. Nevertheless, clear secular trends to earlier peak height velocity in puberty and increased height for age are reliably documented among males and females[57,58]. Gonadal steroid hormones have crucial roles in accelerated growth at puberty, suggesting that the secular growth trends observed in boys, and girls, correspond to secular trends in gonadal activity that stimulate height gains[59].

Causes of accelerated growth and maturation have been intensely investigated. The speed of change rules out genetic bases; rather, the quality of early environments (and in particular nutrition), health and, to a lesser degree, psychosocial stress consistently emerge as major factors[60]. Accelerated growth and timing of puberty have followed the rise of public health measures, improved nutrition and effective maternal–child health care as they have (or have not) spread through populations around the globe[61,62]. Indeed, so tight is this association, that height for age is widely used as a cumulative indicator of environmental conditions[63], including socioeconomic disparities within populations that are commonly reflected by urban–rural differences[64].

Factors that influence the pace and timing of maturation act cumulatively. The greatest sensitivity to resource availability occurs during the period of very rapid growth and development in gestation and the first two years, when poor nutrition and effects of illness have the highest impact. Poor conditions influence body mass and metabolic rate, which determine energetic resources that are available for future growth and reproduction. They also signal poor future conditions and increased mortality risk, acting as cues to adjust life history strategy towards later maturation[65]. For instance, among Filipino men in the Cebu longitudinal health and nutrition survey, weight velocity in the first six months after birth, a proxy for nutritional and growth environment, predicted ages of sexual maturity and initial sexual activity as well as markers of physical embodied capital, such as height at age 22[66]. Moreover, intervention can shift life history outcomes: a controlled study in Guatemalan villages found that supplementation of undernourished infants increased growth and accelerated reproductive milestones, while also improving later function and health, educational and cognitive outcomes, and adult income[67].

Such adjustments in life history are mediated by neuroendocrine processes that bring about trade-offs in resource partitioning. Timing of puberty is the most obvious example, and the psycho-behavioural dimensions of the hormonal changes associated with differences in pubertal timing merit attention. For example, Filipino men with higher infant weight velocity in the above study also had higher testosterone levels as young adults[66]. Testosterone variation among these men was, in turn, associated with mating effort, marital status and parenthood[68]. Increased production of gonadal steroids and adrenal androgens is a consistent, but frequently overlooked, correlate of accelerated maturation, and is associated with good early health and nutrition[10,15]. Therefore, the exposure to steroids of early maturing individuals is both earlier and more intense, adding to the potential impact of secular trends to earlier puberty. Although the reproductive health consequences have been widely discussed[15], current studies that relate pubertal hormones to brain development and behaviour would benefit from exploring those possible correlates in cognitive outcomes through comparative work[69].

Given human sociality and interdependency, psychosocial stress also signals poor environmental quality and informs life history via a neuroendocrine signature that alters stress vulnerability and reactions[70]. The hypothalamic–pituitary–adrenal (HPA) axis regulates stress responses and appears to antagonize HPG activity in healthy adults (stress signals the need to invest in survival, rather than in reproduction). It also partially regulates adrenal androgen production, which undergoes distinctive changes across the life course, including increases that are integral to puberty[10]. Nevertheless, emerging evidence suggests that the conditional coupling of HPA and HPG activity shifts during development[42,68]. Exposure to early life stress may perturb the developmental course of

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

HPG and HPA coupling[71] and contribute to early sexual maturation and accelerated life histories (that is, early maturation and early reproduction) observed among females exposed to high psychosocial stress[72]. Whether there are secular trends in HPA reactivity or HPA–HPG coupling is not known.

To summarize thus far, adolescence reflects the distinctive life history strategy of humans that includes a long childhood and delayed puberty. Moreover, growth and pubertal timing are sensitive to environmental quality (nutrition, health, socioeconomic status and psychosocial stress), displaying developmental reaction norms that are reflective of adaptive capacities for adjustment to current and anticipated resources. Recently, human populations have undergone marked shifts or secular trends in maturational milestones and, by inference, their underlying physiology. These secular trends are uneven within and across populations, and are related to the changing material and social conditions under which development occurs. Such changes illustrate biological embedding in which conditions experienced during development are embodied in both physical and functional outcomes. Next, we discuss the patterns in the cultural configurations of adolescence that shape those conditions.

Cultural constructions of adolescence

Societal changes over the last 200 years have not only altered the conditions for physical development and thus shifted the timing of maturation, but also moved from localized to globalized constructions of adolescence that have redefined the experience of the adolescents. Effects of the rise in schooling and the related globalization of life course schedules are often discussed, but the longer history of diverse, localized cultural formations of adolescence receives less attention. Both are important for the present discussion about relative shifts in biological and cultural timing in life history during the second decade. The cross-cultural comparative literature provides systematic analyses of primary ethnographies to derive general patterns in now mostly historical cultures[73], offering a comprehensive overview with historical depth for considering contemporary formations of adolescence.

Here we draw on descriptive statistics derived from the standard cross-cultural sample (SCCS), comprising ethnographies of 186 preindustrial societies representing optimally independent culture regions (details can be found in refs 74 and 75). Ethnographic material in the SCCS has been rigorously coded on hundreds of variables, including many related to adolescence[76]. Limitations of ethnographic literature for comparative analysis are widely debated[77]. However, our present purposes are not comparative but descriptive, concerned with the distribution of practices that are documented in the record. Although most ethnographies are not focused on adolescence per se, two factors have prompted systematic documentation of this period of life: firstly, the interest of anthropologists in cultural configurations of the life course[78]; secondly, the emphasis on comprehensive description. Consequently, information using nominal coding (present or absent) of defining features of adolescence is sufficient for nearly the entire sample. Establishing the chronology of life events is confounded by the unavailability of chronological age, but as noted previously, in these settings societal responses are cued by the physical signs of puberty. Thus, ethnographic data have been coded for timing of intervention in relation to recognition of initial signs of puberty[79].

Social adolescence appears to have been prevalent, yet both differentiated by gender and widely varied in importance and content among preindustrial societies in the SCCS[80] (Table 1). Unsurprisingly, entry into adolescence in most societies coincided with recognition of puberty, and the transition was often ritually marked in some manner. Over half (56%) held formal initiation ceremonies for either or both sexes, which ritually effected the transition from childhood and instigated a total social transformation encompassing most to all aspects of social life, rather than targeting single aspects (for example, graduation or confirmation)[81]. Initiation ceremonies universally were gender-segregated for initiates, and generally coordinated more closely with physical maturation (onset of menarche) in girls than in boys, who lack a similarly distinct marker. Notably, construction of adolescence as a social condition does not require a specific term for 'adolescence', which was present in only a minority of the relatively small number of societies for which specific information is available. Markers distinguishing adolescents from children need not be linguistic; rather, most societies had distinctive visual markers of adolescence in hairstyle, dress, painting or tattoos, or adornments, whether or not a ritual was held. Visible markers distinguishing adults from adolescents were less common, albeit not rare, suggesting a greater social salience for demarcating childhood from the maturational cascade initiated by puberty[79].

Table 1 | Cultural constructions of adolescence

	Females	Males
Start of adolescence relative to puberty*	Percentage of societies	
Before	17	28
At	82	72
After	1	0
Transition ritualized†	79	68
Formal initiation ritual‡	46	36
Term for adolescent§	41	35
Markers child versus adolescent‖	88	86
Markers adolescent versus adult¶	35	32
Partner choice#	50	57
End of adolescence relative to puberty**		
Early (<2 years)	63	31
Medium (2–4 years)	25	35
Late (>4 year)	12	35
Young-adult stage present††	20	25

Data are from the standard cross-cultural sample ($n = 186$). Puberty is defined as menarche for females (14 if age unspecified); spermarche for males (16 if age unspecified). Number of societies for which the variable could be coded are given in parentheses. Sum of percentages may differ from 100% owing to rounding.
*In societies marking adolescence, onset is based on changes in behaviour or treatment relative to puberty ($n = 182$)[79].
†Ritualized behaviour (formal/public or other) that signifies the childhood–adolescence transition ($n = 130$)[81].
‡Adolescent initiation ceremony ($n = 182$)[99].
§Presence of term for adolescence that is applied to all young people ($n = 41$)[79].
‖Adolescents are differentiated from children by visual markers (dress or ornamentation) ($n = 118$)[79].
¶Adolescents are differentiated from adults by visual markers (dress or ornamentation) ($n = 118$)[79].
#Adolescent has some say as to whether marriage choice occurs in adolescence.
**End of social adolescence, relative to puberty. Girls: early indicates early–mid teens; mid indicates mid–late teens; late indicates late teens–age 20. Boys: early indicates around mid-teens; mid indicate late teens, late indicate early 20s. ($n = 178$)[79].
††Move from adolescence to full adulthood is not direct; a youth phase is included ($n = 168$)[79].

Daily lives of adolescents in these societies were fully integrated into ongoing subsistence activities, and same sex adults commonly dominated their everyday social worlds (66% of SCCS societies for males, 86% for females in 160 and 161 cases, respectively)[79]. Learning occurred informally, mostly through observation and participation with adults, from infancy onward[82], although initiation rituals often included instruction[78]. Therefore, very few societies practiced formal schooling[80], although parents and high-skill mentors might engage in active teaching[83]. Marriage, or socially recognized partnering, usually occurred fairly early, within two years after menarche for girls in their mid to late teens and slightly later for boys, and brought closure to adolescence, with an even distribution as to whether young people had a say in partner choice. Although rapid passage through adolescence to marriage and adulthood was most common, some societies recognized a stage of young adulthood between adolescence and adulthood.

Viewed comprehensively, evidence from preindustrial societies suggests continuities and contrasts with the globalizing contemporary configurations of adolescence. Social adolescence is not a modern innovation, although contemporary formations may be novel; adolescents then and now display distinctive visual markers, and transitions into and out of this period are culturally marked and managed. The radical innovation is compulsory universal schooling[84] (Fig. 3a). Schooling arose along with reliance on chronological age, a shift to peer-dominated settings, and adoption of an education-based wage labour economy, with a concomitant reduction in adolescent productive roles and subsistence activities, and of society-wide initiation rituals[85]. Moreover, schedules

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

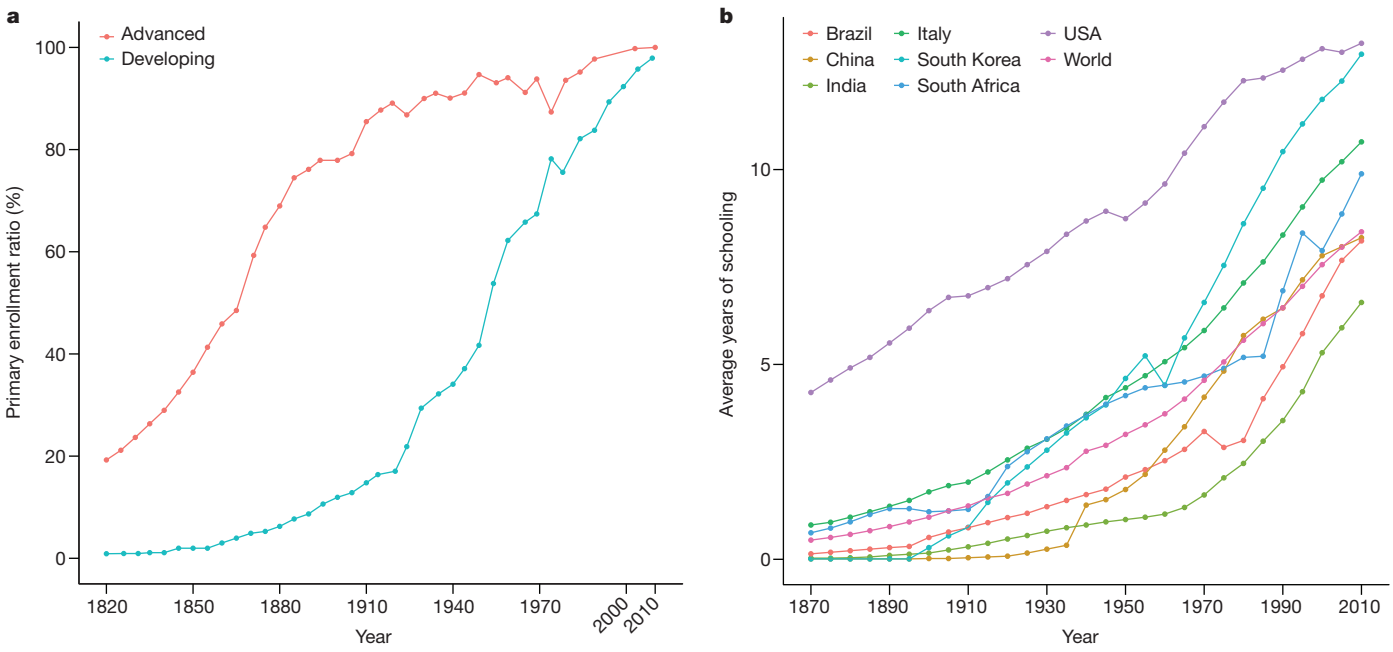


Figure 3 | Trends in education for 1820–2010. a, Trends in primary enrolment ratios (percentage of students enrolled in primary school relative to total primary school-aged population), by 'advanced' and 'developing' economies for 1820–2010. Advanced economies include: Austria, Canada, Japan, Oceania, Turkey, United States and Western Europe. Developing economies include: Asia and the Pacific, excluding Japan; Eastern Europe; Latin America and the Caribbean; the Middle East and North Africa, excluding Turkey; and Sub-Saharan Africa. **b**, Trends in total years of schooling. Drawn from data compiled in ref. 84.

of adolescence have been standardized to progression through school, and prolonged by expanding prerequisites for adulthood that include school completion, gaining employment or an economic foothold, and establishing independence, with a consequent delay in marriage. Unemployment, absent as a category in the preindustrial sample, today affects one in eight youths (over 73 million, three times the number of adults)[86]. Between the years 1970 and 2005, global age at first marriage increased by nearly 6 years in women and 5 years in men[87]. Recognition of a young adult stage (roughly between the age of 18 and 25) has become widespread along with rising thresholds of social criteria for adulthood and uncertainty of attaining them[88].

Why did the global, unprecedented adoption of mass schooling occur? Although abundant social, political and economic rationales motivate mass schooling projects[89], from a life history perspective, formal education represents intensifying investments in functional embodied capital. Changes in survivorship, health and forms of production and prestige prompt facultative adjustments in life history that propel sharply escalating investments in specific forms of functional embodied capital encouraged by schooling[40]. Reduced early mortality and greater life expectancy have increased potential returns from education as labour markets track changing modes of production that reward education- and skill-based embodied capital[5,90]. Accelerated growth and maturation from availability of increased resources for physical embodied capital have converged with pressure for increased investment in functional capacities to greatly expand the window for embodied capital formation at adolescence indexed by an increase in the number of years of schooling[40] (Fig. 3b). The vicissitudes of economic development combined with rising numbers of educated adolescents reinforce competition for employment and a qualifications arms race[91]. Escalating employer demands for labour with novel skills and an education-based capacity for flexibility exacerbate youth unemployment rates, strengthen competition and expand uncertainty and inequality[92,93]. Furthermore, the perception of spiralling demands for investments in embodied capital and related cost of children has been linked to ongoing changes in reproduction and parenting that contribute to widespread fertility declines in the demographic transition[94,95]. These dynamics have remodelled childhood, adolescence and early adulthood.

In summary, cultural formations of adolescence in preindustrial societies differed widely but commonly recognized an adolescent phase that began with signs of puberty and concluded with partnering or marriage. Adolescents usually experienced rites of passage, frequently bore distinctive markers, pursued subsistence tasks in adult-dominated domestic settings, and passed on to marriage and/or adulthood fairly promptly. The rise of compulsory formal education recast adolescent cultural ecology, as defined by chronological age, removed from cooperative domestic productive tasks, installed in competitive peer-dominated settings, and prolonged by educational demands. In life history terms, this considerable shift represents intensifying investment in specific forms of functional embodied capital that education aims to instil. Motivating these changes are structural transformations in labour, livelihood and lifestyle that require specific attitudes, capacities and skills that the young will need for viable futures.

Outlook on strengthening adolescent outcomes

Recent much-discussed shifts in the pace of physical maturation and the cultural formation of adolescence can be explained in life history terms as responses to improved environmental quality on the one hand, and intensifying investments in embodied capital on the other. An apparent discrepancy between biological and cultural processes turns out to represent responses from a shared suite of evolved life history adaptations. Although we describe adolescence from a broad perspective, focusing on population trends that omit variation within and between them, the reality is more complicated. Environmental factors that drive secular trends for physical development also drive within-population variation along urban–rural, socioeconomic, regional or ethnic lines[1]. The well-recognized sources of these differences in physical growth and maturation are targets for policy and intervention from local to global levels that are aimed primarily at infancy and childhood. Yet the distinctive needs of adolescents for promoting their crucial maturational agendas, including growth, reproduction, immune function, cognition and emotion regulation, are only just coming into focus and represent important windows of opportunity to enhance physical embodied capital[2]. As policy and intervention take up this challenge, adolescents as a whole should benefit and inequalities in physical embodied capital should be reduced.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Concurrently, uptake of schooling as a primary means for formation of functional embodied capital aims to foster life chances and mobility to opportunity, yet it also opens new sources of inequality and marginalization that erode the value of this strategy for many[8]. The great gains in average years of schooling (Fig. 3b) still show persistent regional differences, and mask underlying disparities in access (costs and barriers) and quality (teachers, facilities and schooling systems) that generate educational inequality in both years attained and benefits obtained[96]. Furthermore, the relationship between such inequality and gender gaps in education has strengthened with time. Family resources and capacities interact with schooling access and quality to reinforce existing local inequalities in life chances.

Life history analysis of how ecological factors drive cost–benefit tradeoffs that motivate parent and youth investments in embodied capital may inform approaches to these challenges[40]. Such analysis of parental behaviour has shown that inequality drives status competition and intensifies investments in locally valued markers of child quality, such as education or heritable wealth, but the effect collapses when such investments are not possible or will not benefit outcomes[95]. This finding suggests that reductions in educational and economic inequality will increase school participation and performance more effectively than the reverse, as is more commonly assumed[89,96].

In closing, we highlight that context-sensitive maturational processes at adolescence amplify effects from sociocultural ecology and generate special opportunities for biological embedding of culture. This raises the stakes for understanding the ecologies of adolescence, including formal schooling, particularly given current worldwide sociocultural shifts at local to macro levels. As knowledge about ecodevelopmental dynamics continues to accumulate, developmental science is globally expanding, as researchers take up the challenge to investigate a much wider range of human diversity. These exciting advances are stimulating a period of intensifying research and application that promises to vigorously promote adolescent welfare and capability under the diverse local realities that they inhabit within a fluid, globalized world.

Received 14 November 2017; accepted 15 January 2018.

1. Eveleth, P & Tanner, J. *Worldwide Variation in Human Growth* 2nd edn (Cambridge Univ. Press, 1990).
2. Patton, G. C. *et al.* Our future: a *Lancet* commission on adolescent health and wellbeing. *Lancet* **387,** 2423–2478 (2016).
 This comprehensive report thoroughly substantiates the need to invest in adolescents.
3. Worthman, C. M. in *Evolutionary Medicine* (eds Trevathan, W. R. *et al.*) (Oxford Univ. Press, 1999).
4. Stearns, S. C. & Koella, J. C. The evolution of phenotypic plasticity in life-history traits: predictions of reaction norms for age and size at maturity. *Evolution* **40,** 893–913 (1986).
5. Lancaster, J. B. & Kaplan, H. S. in *Human Evolutionary Biology* (ed. Muehlenbein, M. P.) 439–455 (Cambridge Univ. Press, 2010).
6. Byrge, L., Sporns, O. & Smith, L. B. Developmental process emerges from extended brain–body–behavior networks. *Trends Cogn. Sci.* **18,** 395–403 (2014).
 This report offers a dynamic systems account of cognitive–behavioural development.
7. Sawyer, S. M. *et al.* Adolescence: a foundation for future health. *Lancet* **379,** 1630–1640 (2012).
8. Stearns, S. *The Evolution of Life Histories* (Oxford Univ. Press, 1992).
 This foundational text details principles and applications of life history theory.
9. Laland, K., Matthews, B. & Feldman, M. W. An introduction to niche construction theory. *Evol. Ecol.* **30,** 191–202 (2016).
10. Worthman, C. M. in *Hormones, Health, and Behavior: A Socio-ecological and Lifespan Perspective* (eds Panter-Brick, C. & Worthman, C. M.) 47–104 (Cambridge Univ. Press, 1999).
11. Hill, K. & Kaplan, H. Life history traits in humans: theory and empiricial studies. *Annu. Rev. Anthropol.* **28,** 397–430 (1999).
12. Kuzawa, C. W. *et al.* Metabolic costs and evolutionary implications of human brain development. *Proc. Natl Acad. Sci. USA* **111,** 13010–13015 (2014).
13. Livadas, S. & Chrousos, G. P. Control of the onset of puberty. *Curr. Opin. Pediatr.* **28,** 551–558 (2016).
14. Bateson, P. & Gluckman, P. Plasticity and robustness in development and evolution. *Int. J. Epidemiol.* **41,** 219–223 (2012).
15. Jasienska, G., Bribiescas, R. G., Furberg, A. S., Helle, S. & Núñez-de la Mora, A. Human reproduction and health: an evolutionary perspective. *Lancet* **390,** 510–520 (2017).
16. NCD Risk Factor Collaboration (NCD-RisC). A century of trends in adult human height. *eLife* **5,** e13410 (2016).
17. Stout, D. & Hecht, E. E. Evolutionary neuroscience of cumulative culture. *Proc. Natl Acad. Sci. USA* **114,** 7861–7868 (2017).
18. Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The cellular and molecular landscapes of the developing human central nervous system. *Neuron* **89,** 248–268 (2016).
19. Romeo, R. D. Puberty: a period of both organizational and activational effects of steroid hormones on neurobehavioural development. *J. Neuroendocrinol.* **15,** 1185–1192 (2003).
20. Paus, T. *et al.* Sexual dimorphism in the adolescent brain: role of testosterone and androgen receptor in global and local volumes of grey and white matter. *Horm. Behav.* **57,** 63–75 (2010).
21. Koolschijn, P. C. M. & Crone, E. A. Sex differences and structural brain maturation from childhood to early adulthood. *Dev. Cogn. Neurosci.* **5,** 106–118 (2013).
22. Asato, M. R., Terwilliger, R., Woo, J. & Luna, B. White matter development in adolescence: a DTI study. *Cereb. Cortex* **20,** 2122–2131 (2010).
23. Tamnes, C. K. *et al.* Brain maturation in adolescence and young adulthood: regional age-related changes in cortical thickness and white matter volume and microstructure. *Cereb. Cortex* **20,** 534–548 (2010).
24. Kaufmann, T. *et al.* Delayed stabilization and individualization in connectome development are related to psychiatric disorders. *Nat. Neurosci.* **20,** 513–515 (2017).
25. Crone, E. A & Dahl, R. E. Understanding adolescence as a period of social-affective engagement and goal flexibility. *Nat. Rev. Neurosci.* **13,** 636–650 (2012).
 This review tracks important changes in social-affective processing during adolescent neurocognitive development.
26. Galván, A. & McGlennen, K. M. Enhanced striatal sensitivity to aversive reinforcement in adolescents versus adults. *J. Cogn. Neurosci.* **25,** 284–296 (2013).
27. Silk, J. S. *et al.* Increased neural response to peer rejection associated with adolescent depression and pubertal development. *Soc. Cogn. Affect. Neurosci.* **9,** 1798–1807 (2014).
28. Gunther Moor, B. *et al.* Social exclusion and punishment of excluders: neural correlates and developmental trajectories. *Neuroimage* **59,** 708–717 (2012).
29. Chein, J., Albert, D., O'Brien, L., Uckert, K. & Steinberg, L. Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Dev. Sci.* **14,** F1–F10 (2011).
30. van Duijvenvoorde, A. C. K., Zanolie, K., Rombouts, S. A. R. B., Raijmakers, M. E. J. & Crone, E. A. Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *J. Neurosci.* **28,** 9495–9503 (2008).
31. Spielberg, J. M., Olino, T. M., Forbes, E. E. & Dahl, R. E. Exciting fear in adolescence: does pubertal development alter threat processing? *Dev. Cogn. Neurosci.* **8,** 86–95 (2014).
32. Rose, A. J. & Rudolph, K. D. A review of sex differences in peer relationship processes: potential trade-offs for the emotional and behavioral development of girls and boys. *Psychol. Bull.* **132,** 98–131 (2006).
33. Blakemore, S.-J. & Mills, K. L. Is adolescence a sensitive period for sociocultural processing? *Annu. Rev. Psychol.* **65,** 187–207 (2014).
 This review compiles evidence of social contextual effects on adolescent brain maturation.
34. Boyd, R., Richerson, P. J. & Henrich, J. The cultural niche: why social learning is essential for human adaptation. *Proc. Natl Acad. Sci. USA* **108** (Suppl 2), 10918–10925 (2011).
 This analysis documents the crucial role of social learning in human adaptation.
35. Chudek, M., Brosseau-Liard, P., Birch, S. & Henrich, J. in *Navigating the Social World: What Infants, Children, and other Species can teach Us* (eds Banaji, M. R. & Gelman, S. A.) 181–185 (Oxford Univ. Press, 2013).
36. Henrich, J. *et al. Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (Oxford Univ. Press, 2004).
37. Creanza, N., Kolodny, O. & Feldman, M. W. Cultural evolutionary theory: how culture evolves and why it matters. *Proc. Natl Acad. Sci. USA* **114,** 7782–7789 (2017).
38. Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A. & Henrich, J. Two ways to the top: evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *J. Pers. Soc. Psychol.* **104,** 103–125 (2013).
39. Whiting, B. B & Edwards, C. P. *Children of Different Worlds: The Formation of Social Behavior* (Harvard Univ. Press, 1988).
40. Kaplan, H., Lancaster, J. & Robson, A. Embodied capital and the evolutionary economics of the human life span. *Popul. Dev. Rev.* **29,** 152–182 (2003).
 This article uses the concept of embodied capital in an evolutionary account of life history.
41. Dickemann, M. in *Evolutionary Biology and Human Social Behavior* (eds Chagnon, N. & Irons, W.) 321–367 (Duxbury, 1979).
42. Shirtcliff, E. A. *et al.* A dual-axis approach to understanding neuroendocrine development. *Dev. Psychobiol.* **57,** 643–653 (2015).
43. Biro, F. M. *et al.* Pubertal assessment method and baseline characteristics in a mixed longitudinal study of girls. *Pediatrics* **126,** e583–e590 (2010).
44. Abbassi, V. Growth and normal puberty. *Pediatrics* **102,** 507–511 (1998).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

45. Juul, A., Magnusdottir, S., Scheike, T., Prytz, S. & Skakkebaek, N. E. Age at voice break in Danish boys: effects of pre-pubertal body mass index and secular trend. *Int. J. Androl.* **30,** 537–542 (2007).
46. Macmahon, B. *et al.* Age at menarche, probability of ovulation and breast cancer risk. *Int. J. Cancer* **29,** 13–16 (1982).
47. Tanner, J. M. Trend towards earlier menarche in London, Olso, Copenhagen, the Netherlands and Hungary. *Nature* **243,** 95–96 (1973).
48. Atay, Z., Turan, S., Guran, T., Furman, A. & Bereket, A. Puberty and influencing factors in schoolgirls living in Istanbul: end of the secular trend? *Pediatrics* **128,** e40–e45 (2011).
49. Jones, L. L., Griffiths, P. L., Norris, S. A., Pettifor, J. M. & Cameron, N. Age at menarche and the evidence for a positive secular trend in urban South Africa. *Am. J. Hum. Biol.* **21,** 130–132 (2009).
50. Sørensen, K. *et al.* Recent secular trends in pubertal timing: implications for evaluation and diagnosis of precocious puberty. *Horm. Res. Paediatr.* **77,** 137–145 (2012).
51. Parent, A.-S. *et al.* The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr. Rev.* **24,** 668–693 (2003).
52. Mouritsen, A. *et al.* Hypothesis: exposure to endocrine-disrupting chemicals may interfere with timing of puberty. *Int. J. Androl.* **33,** 346–359 (2010).
53. Goldstein, J. R. A secular trend toward earlier male sexual maturity: evidence from shifting ages of male young adult mortality. *PLoS ONE* **6,** e14826 (2011).
54. Tomova, A., Lalabonova, C., Robeva, R. N. & Kumanov, P. T. Timing of pubertal maturation according to the age at first conscious ejaculation. *Andrologia* **43,** 163–166 (2011).
55. Sørensen, K., Aksglaede, L., Petersen, J. H. & Juul, A. Recent changes in pubertal timing in healthy Danish boys: associations with body mass index. *J. Clin. Endocrinol. Metab.* **95,** 263–270 (2010).
56. Papadimitriou, A. *et al.* Pubertal maturation of contemporary Greek boys: no evidence of a secular trend. *J. Adolesc. Health* **49,** 434–436 (2011).
57. Aksglaede, L., Olsen, L. W., Sørensen, T. I. & Juul, A. Forty years trends in timing of pubertal growth spurt in 157,000 Danish school children. *PLoS ONE* **3,** e2728 (2008).
58. Ali, M. A., Lestrel, P. E. & Ohtsuki, F. Secular trends for takeoff and maximum adolescent growth for eight decades of Japanese cohort data. *Am. J. Hum. Biol.* **12,** 702–712 (2000).
59. Perry, R. J., Farquharson, C. & Ahmed, S. F. The role of sex steroids in controlling pubertal growth. *Clin. Endocrinol.* **68,** 4–15 (2008).
60. Perkins, J. M., Subramanian, S. V., Davey Smith, G. & Özaltin, E. Adult height, nutrition, and population health. *Nutr. Rev.* **74,** 149–165 (2016).
61. Batty, G. D. *et al.* Height, wealth, and health: an overview with new data from three longitudinal studies. *Econ. Hum. Biol.* **7,** 137–152 (2009).
62. Moradi, A. Nutritional status and economic development in sub-Saharan Africa, 1950–1980. *Econ. Hum. Biol.* **8,** 16–29 (2010).
63. Steckel, R. H. Heights and human welfare: recent developments and new directions. *Explor. Econ. Hist.* **46,** 1–23 (2009).
64. Subramanian, S. V., Özaltin, E. & Finlay, J. E. Height of nations: a socioeconomic analysis of cohort differences and patterns among women in 54 low- to middle-income countries. *PLoS ONE* **6,** e18962 (2011).
65. Kuzawa, C. W. Developmental origins of life history: growth, productivity, and reproduction. *Am. J. Hum. Biol.* **19,** 654–661 (2007).
66. Kuzawa, C. W., McDade, T. W., Adair, L. S. & Lee, N. Rapid weight gain after birth predicts life history and reproductive strategy in Filipino males. *Proc. Natl Acad. Sci. USA* **107,** 16800–16805 (2010).
67. Martorell, R. Improved nutrition in the first 1000 days and adult human capital and health. *Am. J. Hum. Biol.* **29,** e22952 (2017).
68. Gettler, L. T., McDade, T. W. & Kuzawa, C. W. Cortisol and testosterone in Filipino young adult men: evidence for co-regulation of both hormones by fatherhood and relationship status. *Am. J. Hum. Biol.* **23,** 609–620 (2011).
69. Schulz, K. M & Sisk, C. L. The organizing actions of adolescent gonadal steroid hormones on brain and behavioral development. *Neurosci. Biobehav. Rev.* **70,** 148–158 (2016).
This review examines sex-specific effects of gonadal steroid hormones on adolescent brain and behaviour.
70. De Bellis, M. D. & Zisk, A. The biological effects of childhood trauma. *Child Adolesc. Psychiatr. Clin. N. Am.* **23,** 185–222 (2014).
71. Ruttle, P. L., Shirtcliff, E. A., Armstrong, J. M., Klein, M. H. & Essex, M. J. Neuroendocrine coupling across adolescence and the longitudinal influence of early life stress. *Dev. Psychobiol.* **57,** 688–704 (2015).
72. Belsky, J., Schlomer, G. L. & Ellis, B. J. Beyond cumulative risk: distinguishing harshness and unpredictability as determinants of parenting and early life history strategy. *Dev. Psychol.* **48,** 662–673 (2012).
73. Ember, C. R. Using the HRAF collection of ethnography in conjunction with the standard cross-cultural sample and the ethnographic atlas. *Cross Cult. Res.* **41,** 396–427 (2007).
74. Murdock, G. P. & Murdock, D. R. Standard cross-cultural sample. *Ethnology* **8,** 329–396 (1969).
75. White, D. R. Focused ethnographic bibliography: standard cross-cultural sample. *Cross Cult. Res.* **23,** 1–145 (1989).
76. Barry, H & Schlegel, A. *Cross-Cultural Samples and Codes* (Univ. Pittsburgh Press, 1980).
77. Mace, R. & Pagel, M. The comparative method in anthropology. *Curr. Anthropol.* **35,** 549–564 (1994).
78. van Gennep, A. *The Rites of Passage* (Univ. Chicago Press, 1960).
79. Schlegel, A. & Barry, H. *Adolescence: An Anthropological Inquiry* (Free Press, 1991).
On the basis of extensive ethnographic data, this work offers a detailed overview of cultural constructions of adolescence.
80. Schlegel, A. A cross-cultural approach to adolescence. *Ethos* **23,** 15–32 (1995).
81. Schlegel, A. & Barry, H. The evolutionary significance of adolescent initiation ceremonies. *Am. Ethnol.* **7,** 696–715 (1980).
82. Paradise, R. & Rogoff, B. Side by side: learning by observing and pitching in. *Ethos* **37,** 102–138 (2009).
83. Kline, M. A., Boyd, R. & Henrich, J. Teaching and the life history of cultural transmission in Fijian villages. *Hum. Nat.* **24,** 351–374 (2013).
84. Lee, J. W. & Lee, H. Human capital in the long run. *J. Dev. Econ.* **122,** 147–169 (2016).
85. Greenfield, P. M. Linking social change and developmental change: shifting pathways of human development. *Dev. Psychol.* **45,** 401–418 (2009).
86. International Labor Organization. *World Employment and Social Outlook.* http://www.ilo.org/global/research/global-reports/weso/2017/lang--en/index.htm (International Labor Office, 2017).
87. United Nations. *World Marriage Data 2012.* http://www.un.org/esa/population/publications/WMD2012/MainFrame.html (Department of Economic and Social Affairs, Population Division, 2013).
88. Lloyd, C. B. (ed.) *Growing up Global: The Changing Transitions to Adulthood in Developing Countries* (National Academies Press, 2006).
89. Hannum, E & Buchman, C. *The Consequences of Global Educational Expansion* (American Academy of Arts and Sciences, 2003).
90. Gakidou, E., Cowling, K., Lozano, R. & Murray, C. J. L. Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: a systematic analysis. *Lancet* **376,** 959–974 (2010).
91. Worthman, C. M. Inside-out or outside-in? Global development theory, policy, and youth. *Ethos* **39,** 432–451 (2011).
92. Johnson-Hanks, J. in *Ethnographies of Youth and Temporality* (eds Frederiksen, M. D. & Dalsgård, A. L.) 23–56 (Temple Univ. Press, 2014).
93. International Labor Organization. *Global Employment Trends for Youth 2013: A Generation at Risk.* http://www.ilo.org/global/research/global-reports/youth/2013/lang--en/index.htm (International Labor Office, 2013).
94. Bongaarts, J. Human population growth and the demographic transition. *Phil. Trans. R. Soc. B* **364,** 2985–2990 (2009).
95. Shenk, M. K., Kaplan, H. S. & Hooper, P. L. Status competition, inequality, and fertility: implications for the demographic transition. *Phil. Trans. R. Soc. B* **371,** 20150150 (2016).
96. Thomas, V., Wang, Y & Fan, X. *Measuring education inequality: Gini coefficients of education.* Policy Research Working Paper 2525 http://documents.worldbank.org/curated/en/361761468761690314/Measuring-education-inequality-Gini-coefficients-of-education (World Bank, 2001).
97. Stefan, N., Häring, H. U., Hu, F. B. & Schulze, M. B. Divergent associations of height with cardiometabolic disease and cancer: epidemiology, pathophysiology, and global implications. *Lancet Diabetes Endocrinol.* **4,** 457–467 (2016).
98. Song, Y. *et al.* Trends of age at menarche and association with body mass index in Chinese school-aged girls, 1985–2010. *J. Pediatr.* **165,** 1172–1177 (2014).
99. Schlegel, A. & Barry, H. Adolescent initiation ceremonies: a cross-cultural code. *Ethnology* **18,** 199 (1979).

Author Contributions K.T. assembled population data; C.M.W. filtered ethnographic data; C.M.W. and K.T. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to C.M.W. (worthman@emory.edu).

Reviewer Information *Nature* thanks S. Spyrou and D. F. Lancy for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ANALYSIS

OPEN

doi:10.1038/nature25759

Adolescence and the next generation

George C. Patton[1,2,3], Craig A. Olsson[1,2,3,4], Vegard Skirbekk[5,6], Richard Saffery[1,2], Mary E. Wlodek[7], Peter S. Azzopardi[1,2,8,9], Marcin Stonawski[10,11], Bruce Rasmussen[12], Elizabeth Spry[2,3,4], Kate Francis[2,3], Zulfiqar A. Bhutta[13,14], Nicholas J. Kassebaum[15,16], Ali H. Mokdad[15], Christopher J. L. Murray[15], Andrew M. Prentice[17,18], Nicola Reavley[19], Peter Sheehan[12], Kim Sweeny[12], Russell M. Viner[20] & Susan M. Sawyer[1,2,3]

Adolescent growth and social development shape the early development of offspring from preconception through to the post-partum period through distinct processes in males and females. At a time of great change in the forces shaping adolescence, including the timing of parenthood, investments in today's adolescents, the largest cohort in human history, will yield great dividends for future generations.

Global megatrends are reshaping health and human development almost everywhere[1,2]. Rapid economic, technological, social and demographic changes have brought reductions in infectious diseases, infant and maternal mortality and refocused attention on the non-communicable diseases of later life[3]. These forces are also reshaping health and development across the second and third decades. In preindustrial societies, the time between the end of childhood, which is signalled in girls by the late pubertal event of menarche (onset of menstruation) and spermarche (first ejaculation) or breaking of the voice in boys, and transition to adulthood, typically marked by parenthood, was generally around two years in girls and four in boys[4,5]. Improved nutrition and fewer infectious diseases in childhood, have been accompanied by a fall of around four years in the age at puberty to between 12 and 13 years[6], a fall that has been rapid in current middle- and lower-income countries[7,8]. Over the same period, an even bigger upward shift has occurred in the timing of parenthood, the result of extended education, changing social norms around marriage and parenthood, and the availability of effective contraception.

Puberty initiates a phase of growth and maturation of the reproductive, musculoskeletal, neurodevelopmental, endocrine, metabolic, immune and cardio-metabolic systems, that extends into the third decade[9,10]. For this reason, adolescence can be considered a sensitive phase, during which the quality of the physical, nutritional and social environments may change trajectories of health and development into later life[11]. Given the concept that growth continues into the twenties, together with the delays in adopting adult roles, the idea has been proposed that adolescence might best be considered as ranging from 10 to 24 years[12]. From this perspective, adolescence occupies a greater proportion of the life course with greater relevance for human development than ever before[9]. An extended adolescence creates an opportunity for this generation to acquire greater assets and capabilities. Equally, adolescence has been accompanied by shifts in the social milieu of development, with the emergence of distinct youth cultures, greater media and peer engagement, and marketing to future consumers, in turn shifting patterns of health and health risk[13]. For example, the risk of sexually transmitted infections increases with multiple sexual partners before marriage; earlier initiation of substance use is associated with greater risk of later substance-use disorders; and a reduction in physical activity, alongside changes in diet, is associated with higher rates of obesity[11,14].

In contrast to the recognition of the significance of adolescence for later adult health, its relevance for the next generation has received less attention[15]. This is surprising, given that adolescents are the next generation to become a parent and we have known for decades that preconception maternal nutrition (for example, folate deficiency)[16] and infectious diseases[17,18] affect the early life health and development of offspring. Indeed, a failure to consider influences on growth during early life that emerge in adolescence before conception may explain why antenatal interventions have too often only led to small gains[10,19].

Here we explore the range of adolescent processes—molecular, physiological, behavioural and sociocultural—that may affect the early growth, health and development of the next generation[19–21] (Fig. 1). Firstly, we consider time trends in the age of first parenthood and overall fertility across different country groups classified by level of economic development. Secondly, we explore the range of potentially modifiable processes that start during adolescence that may affect the early growth of offspring. Finally, we consider the shifting social determinants of health to illustrate the benefits for the next generation by investing in adolescents as the parents of tomorrow.

The changing age of parenthood

Lower rates of giving birth during adolescence have led to a later transition to parenthood in most places. In all country income groups, birth rates in 15–19-year-old girls continue to fall with the most marked shifts in today's lower middle-income countries where rates in 2030 are anticipated to be a third of those in 1970 (Extended Data Fig. 1a). Marked falls are also taking place across all country groups for young women aged 20–24, with 2030 birth rates in today's high-income countries also predicted to be a third of the rates in 1970. As a result, adolescent parenthood

[1]The University of Melbourne, Department of Paediatrics, Faculty of Medicine, Dentistry and Health Sciences, Parkville, Victoria 3010, Australia. [2]Murdoch Children's Research Institute, Parkville, Victoria 3052, Australia. [3]Centre for Adolescent Health, Royal Children's Hospital, Parkville, Victoria 3052, Australia. [4]Deakin University Geelong, Centre for Social and Early Emotional Development, School of Psychology, Faculty of Health, Geelong, Victoria 3220, Australia. [5]Centre for Fertility and Health, Norwegian Institute of Public Health, Nydalen, Oslo 0403, Norway. [6]Columbia University, New York, New York 10032, USA. [7]The University of Melbourne, Department of Physiology, Parkville, Victoria 3010, Australia. [8]Maternal and Child Health Program, International Development Discipline, Burnet Institute, Melbourne, Victoria 3004, Australia. [9]Wardliparingga Aboriginal Research Unit, South Australian Health and Medical Research Institute, Adelaide, South Australia 5000, Australia. [10]Department of Demography, Cracow University of Economics, Cracow 31-510, Poland. [11]European Commission, Joint Research Centre, Centre for Advanced Studies, Ispra, Varese 21027, Italy. [12]Victoria Institute of Strategic Economic Studies, Victoria University, Melbourne, Victoria 3000, Australia. [13]SickKids Centre for Global Child Health, Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada. [14]Centre of Excellence in Women and Child Health, Aga Khan University, Karachi 74800, Pakistan. [15]Institute for Health Metrics and Evaluation, University of Washington, Seattle, Washington 98121, USA. [16]Division of Pediatric Anesthesiology & Pain Medicine, Seattle Children's Hospital, Seattle, Washington 98105, USA. [17]MRC Unit The Gambia, Fajara, Gambia. [18]MRC International Nutrition Group, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK. [19]The University of Melbourne, Melbourne School of Population and Global Health, Parkville, Victoria 3010, Australia. [20]UCL Institute of Child Health, University College London, London WC1N 1EH, UK.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

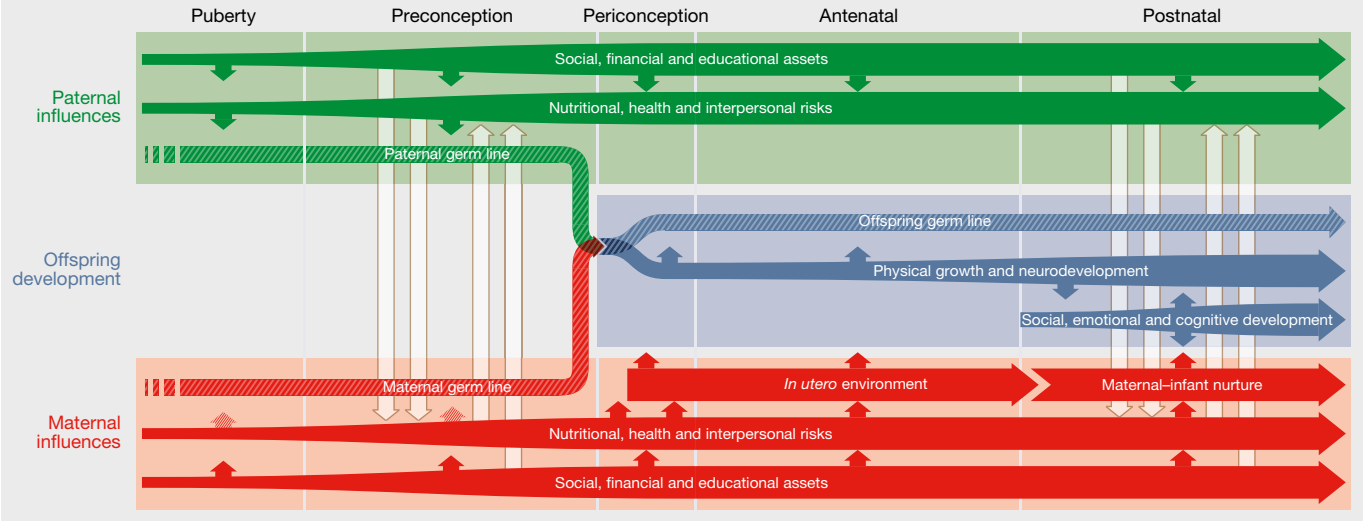


Figure 1 | Adolescence and the next generation: a model of processes underpinning intergenerational transmission. Puberty marks a transition to adolescence and a life phase during which girls and boys acquire resources that are essential for becoming parents of the next generation. It also marks the beginning of reproductive life with a transition to functional gamete production. The preconceptional phase (that is, adolescence) varies markedly in length carrying implications for the acquisition of the social, financial and educational assets and nutritional, health and interpersonal risks that underlie intergenerational processes. The three months before conception is a time of male and female gamete maturation when parental exposures, including nutrition, obesity, substance use, stress, endocrine disruptors and physical activity may influence gamete structure and function. Periconception includes fertilization of the maternal and paternal gametes as well as the zygote and embryonic phases that are sensitive to the maternal nutritional and hormonal environment. There continue to be direct maternal effects antenatally mediated through the *in utero* environment and postnatally through nutrition (for example, breastfeeding) and the maternal–infant relationship. Direct paternal influences grow in the postnatal phase through the paternal–infant relationship and potentially through risk exposures, such as paternal tobacco use. Maternal and paternal health, behaviour as well as social and economic circumstances continue to have an indirect effect on offspring development in both antenatal and postnatal phases.

is increasingly concentrated in specific geographic regions (South Asia, Latin-America and sub-Saharan Africa) and disadvantaged groups within countries[22].

Decreases in birth rates during adolescence have led to parallel shifts in the duration of both adolescence and the preconception phase of reproductive life, defined as the time from reproductive maturity to first parenthood (Extended Data Fig. 1b). In high-income countries, only a quarter of girls now have a child by 25 years of age. As a result the gap between reproductive maturity, marked by menarche (on average between 12 and 13 years), and first parenthood is generally over a decade and often extends up to two. By contrast, over 80% of girls growing up in low-income countries become a parent by the age of 25, and 30% by the age of 18 years (Table 1). These differences have major implications for the acquisition of both assets and risks during the preconception window (Fig. 1).

Data on the timing of parenthood for males are scarcer and generally of poorer quality. France and Norway are among the few countries with data[23]. French men typically have children at around the age of 33 years, three years later than women. Males are also coming to parenthood later in life; 40% of Norwegian men born in the 1950s had become a parent by the age of 25 compared to less than 20% for those born in the 1970s, and just over 10% of those born in the 1990s. Therefore, the gap between reproductive maturity and parenthood has also lengthened for boys, a result of effective contraception, extended education, higher costs to establish a family and changing social norms around family formation[24].

Table 1 | Percentage of mothers giving birth aged 15–24 years

	15–24-year-old females	Percentage of females who gave birth before age:				Countries with data
	n	15	18	20	25	
Low income	63,744,000	5.5	30.1	53.2	84.1	28
Lower middle income	265,731,000	4.1	23.0	41.9	74.9	42
Upper middle income	179,542,000	1.7	15.7	32.6	65.7	26
High income	69,695,000	0.2	2.0	7.1	32.1	6

Percentage of women aged 25–29 who reported giving birth before specific ages from 15 to 24 years in household surveys over the past twenty years. The numbers of 15–24-year-old females are from ref. 140.

Early parenthood and its consequences

A longer adolescence leads to healthier growth, particularly for girls without the competing nutritional demands of early pregnancy, and greater opportunities for education and entry into the workforce[25]. For these reasons, age at first birth is a powerful determinant of long-term parental economic capacity, including income and housing[26]. Conversely adolescent parenthood predicts both poor maternal and infant health. In over 120,000 pregnancies in girls aged 10–24 in the Multicountry Survey of Maternal and Child Health of the World Health Organization (WHO), those giving birth before 15 years of age were at higher risk of eclampsia and puerperal infection[27]. Their babies had higher rates of low birth weight, preterm delivery and severe neonatal conditions[27]. Similarly, in pooled analyses of 19,403 offspring from birth cohorts in Brazil, Guatemala, India, the Philippines and South Africa, giving birth before the age of 19 predicted low birth weight, preterm birth and stunting of offspring and these associations were strongest in the youngest mothers[28].

Nutritional needs increase markedly with rapid pubertal growth, making younger adolescents and their offspring vulnerable to undernutrition[29]. Studies of adolescents exposed to famine provide compelling evidence of the persisting effects on adult height of severe undernutrition during adolescence[30]. Equally, follow-up of the offspring of parents exposed to the Chinese (1959–1961) and Cambodian (1975–1979) famines suggest intergenerational risks, with rates of stunting greatest in those whose parents were exposed during adolescence[31,32]. Gestational undernutrition is particularly prominent in adolescence with maternal–fetal competition for energy and nutrients that remain essential for a girl's own continued growth[33]. It is associated with increased risk of neonatal mortality, preterm birth, small for gestational age babies, and low birth weight[34]. The effects extend to the postnatal period with compromised antenatal breast development reducing breast milk quality and quantity[35] and the establishment of attachment bonds[36,37].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

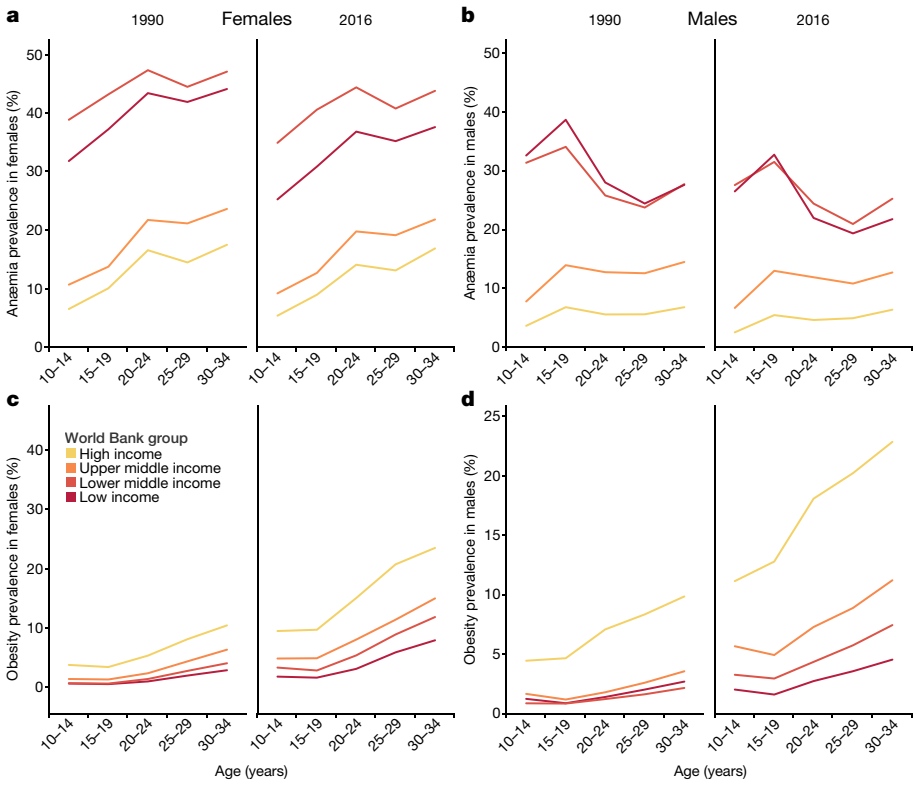


Figure 2 | Nutrition risks across the years of transition to first pregnancy for 1990 and 2016. a, **b**, Prevalence of anaemia in females (**a**) and males (**b**). Prevalence of obesity in females (**c**) and males (**d**).

Iron and other micronutrient requirements increase sharply at puberty, particularly for girls with the onset of menstruation, so that in younger adolescents anaemia is the leading cause of global disease burden[14]. Adolescent anaemia is commonly a result of iron and other micronutrient deficiencies and is therefore a useful indicator of nutritional status[38,39]. Figure 2a and Extended Data Figure 2 show rates of anaemia across the years of transition to first parenthood. Since 1990, rates of anaemia in low- and lower middle-income countries have fallen but only modestly. Consistent with rapid growth and the onset of menstruation, anaemia rates rise for young women from adolescence into the early twenties and then stabilize through the remaining reproductive years. It indicates that nutritional deficiencies remain common in young women in the transition to parenthood and policies to date have failed to address these major, modifiable determinants of early growth in the next generation.

In high-income countries, maternal and offspring outcomes are also poorer in those giving birth before 15 years of age with greater infant death, stillbirth, intrauterine growth restriction and preterm birth[40]. Teenage childbearing is often unplanned, and is associated with partnership instability, a greater likelihood of single parenthood, greater poverty and less parental education contributing to poorer child outcomes[41,42]. Even small delays of 6–18 months in the transition to motherhood independently predict offspring test scores for reading and mathematics that are in turn predictive of educational attainment and eventual earnings[43].

Antenatal actions to promote maternal and fetal health generally occur after pregnancy recognition or a first antenatal visit. In high-income countries, such as the United States, women are generally unaware of pregnancy until six weeks of gestation or longer if the pregnancy was unintended[44]. Initiation of antenatal care is typically around three weeks later[45]. In resource poor settings, pregnancy confirmation and a first antenatal visit are generally much later. Less than half of pregnant women in sub-Saharan Africa have an antenatal visit in the first trimester, and rates are lower in adolescents[46]. Although undernutrition has been a major policy focus in maternal health[47], antenatal interventions with maternal protein-energy or micronutrient supplements have generally brought minimal changes to birth weight[10]. One probable reason is that interventions are taking place too late in pregnancy[48].

Intergenerational risk processes

A range of parental adaptations have the potential to influence the growth and development of the next generation. These biological, interpersonal and social processes are outlined in the following sections.

Mechanisms involving parental gametes

A possibility that the male germline captures information from a changing environment to pass on to subsequent generations has generated attention more recently[49]. It challenges an assumption that the sole function of gametes is to deliver half each of the maternal and paternal genomes to a common zygote. It was believed that this process did not include the transfer of environmental information. However, alternative theories, starting with Lamarck, and including Darwin in his theory of pangenesis, have suggested that some adaptations may be transmitted to the next generation. Recent studies have begun to describe these processes, revealing an interplay between environmental exposures and the parental reproductive milieu (Fig. 1). Potential mechanisms involve not only the protein and RNA cargo contained within gametes, but also factors, which are external to gametes, that are capable of influencing fertilization and early embryonic development. Gamete DNA also has a distinct, and modifiable, epigenetic profile, that is sensitive to environmental exposures, with some changes that are maintained during embryonic development, thus carrying information from one generation to the next[50].

For both males and females, puberty initiates functional gamete production, a process that continues throughout adult reproductive life. There are likely to be sex differences in the timing and duration of sensitive exposure windows, given differences in gamete maturation. Gametogenesis begins early *in utero* with the specification of primordial germ cells. In males, this process is halted before meiosis, but in females, the process proceeds to the first meiotic division, where it then arrests until puberty. Thus, all primary oocytes produced throughout a female's lifetime exist before birth, or shortly thereafter. By contrast, the process of spermatogenesis commences at puberty. Sperm maturation takes between 70 and 100 days in humans, predominantly in the seminiferous tubules that are in constant contact with supporting Sertoli cells. This is followed by a 1–2 week transitional period in the epididymis, during which time

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

proteins and a small amount of RNA are selectively shuttled into the developing sperm[51,52].

The paternal preconception environment modulates the RNA content of the developing sperm, with possible effects on post-fertilization development and offspring phenotypes[53]. In animal studies, paternal exposures to high-fat and high-sugar diets produce metabolic disturbances in offspring[54], and both stress and exercise affect stress responses of the offspring[55–57]. Animal and some human studies have found effects of preconceptional tobacco, alcohol and illicit drug use on gamete epigenomes, with diverse effects on offspring development[58,59]. Although there are fewer human studies, alcohol consumption of more than two standard drinks per day is linked to morphological changes in sperm[60]. One report of a very large study has found effects of heavy (more than five standard drinks per day) paternal but not maternal preconception drinking on offspring head circumference and risk of microcephaly[61]. Paternal obesity generates risks of metabolic disturbance and obesity in offspring through the gamete epigenome, but diet and exercise appear to modify these risks[62,63]. In humans, bariatric surgery reverses obesity-induced epigenetic changes in spermatozoa[64]. Whether these paternal processes are 'anticipatory adaptive responses' to a likely postnatal environment or have other functions is currently under debate[65–67].

Maternal gametes (oocytes) also have a unique epigenetic profile, as well as a cargo of RNA species and proteins that are mostly acquired over the four months of maturation before ovulation[68]. Both animal and human studies have shown that the preconception environment may alter oocyte maturation[69] with maternal obesity affecting the metabolism of the developing ova and early offspring growth[70,71]. Because mitochondrial DNA is only passed on through the maternal line, oocyte mitochondrial DNA may provide a further sex-specific intergenerational mechanism. In a mouse study, a high-fat and high-sugar diet beginning before conception altered insulin signalling in the skeletal muscle of the offspring due to mitochondrial dysfunction, an effect that passed through the maternal germ line to the third generation[72].

Periconceptional mechanisms

The time between fertilization and embryo implantation into the uterine decidua is marked by the emergence of distinct cell lineages, beginning with the specification of embryonic and extra-embryonic cells. There is extensive remodelling of parental gametic epigenetic profiles, a process that differs depending on the parent of origin. This re-organization initiates a shift from two distinct gametic epigenomes to a single embryonic epigenome with a capacity to form any cell type (see Fig. 1). This embryonic phase consists of two major epigenetic remodelling events: the first occurs immediately after fertilization; and the second induces the reestablishment of totipotency in primordial germ cells. These cells are the precursors of all offspring gametes, which is one of the reasons why periconceptional processes have implications for more than one generation[73].

The epigenomic reorganization of early embryogenesis is sensitive to the maternal environment, with effects of protein-energy and micronutrient deficiencies as well as over-nutrition[74,75]. Observational studies, such as the Dutch Hunger Winter study, have implicated periconception as a sensitive time for epigenetic programming[76]. Of note are studies of metastable epialleles, epigenetically labile regions of the genome that are subject to large changes in response to environmental influences during very early embryogenesis. These regions were first described in mice, but subsequent human studies in Gambia have demonstrated epigenetic variation at metastable epialleles that are linked to periconceptional maternal nutrition[77], findings that have later been replicated in a rural Bangladeshi cohort[78]. The methylation status of metastable epialleles has been linked to later obesity and metabolic function[79].

Implantation is associated with further sensitivity to environmental influence. It initiates a functional interaction between the blastocyst (early embryo) and endometrium. The uterine microenvironment allows embryo–maternal crosstalk both pre- and post-implantation. Extracellular vesicles produced by the endometrium have been shown to be mediators of maternal–embryo communication through their protein and RNA cargoes[80]. Secreted proteins from both the embryo (for example, human chorionic gonadotrophin from the trophectoderm) and endometrium (for example, cytokines) guide successful implantation[81]. These processes are sensitive to maternal hormonal status and environmental influences, including stress and nutrition[82].

Persistence of adolescent assets and risks

Assets derived from education, financial resources, family, social and cultural capital that are acquired during adolescence are essential for effective parenting and ultimately the growth and development of the next generation[83] (see Fig. 1). Similarly, risks related to undernutrition, obesity, substance abuse and mental disorders become prominent following puberty and tend to persist through the transition to parenthood even when this occurs decades later[84].

Stress and mental disorders. Around one in four mothers experience symptoms of depression and anxiety in the perinatal period[85]. In contrast to earlier views that perinatal depression is a discrete disorder limited to the post-partum period, perinatal maternal depression more commonly seems to be a continuation of pre-pregnancy mental health problems into pregnancy and the post-partum period. A recent Australian study found that 86% of mothers with high perinatal depressive symptoms had a similar history before conception, predominantly dating back to adolescence[86]. Antenatal and postnatal exposures independently affect child cognitive and emotional development[85,87]. In low-resource settings, the effects extend to childhood stunting and physical illness with antenatal depressive symptoms predicting higher rates of pre-term birth, failure to thrive *in utero* and low birth weight[88,89]. Animal studies of maternal antenatal stress have shown effects on various hormonal and chemical mediators (glucocorticoids, oxygen and glucose) that are associated with reduced fetal weight, gender-specific metabolic changes in the offspring and long-term adverse metabolic and renal effects[90,91]. Persistence of preconception mental health risks into the postpartum affects mother–infant bonding with a greater likelihood of maternal over-intrusiveness, emotional withdrawal and failure to sensitively engage[92].

Given recent secular trends to earlier onset, adolescent mental disorders have become more important as an intergenerational risk process[93]. In high-income settings, over half of young women have an episode of a common mental disorder before becoming a parent[94]. Figure 3a and Extended Data Figure 3 show prevalence estimates for DSM (Diagnostic and Statistical Manual of Mental Disorders) depressive disorders (major depression and dysthymia) in females and males aged 10–35 years, drawn from the Global Burden of Disease (2016) study[95]. Rates increase through to the mid-twenties and then plateau so that in high-income countries, women typically become parents at a time of high risk.

Substance use. Experimentation with tobacco, alcohol and other drugs typically begins in adolescence, with escalation to higher risk use and dependence in early adulthood[96]. Recent trends have generally been to earlier and heavier use with growing similarities in use in females and males[97]. Figure 3b, c and Extended Data Figures 4, 5 show patterns of daily smoking and alcohol-use disorders in females and males across the years of transition to parenthood. Daily smoking rates rise steeply during the period of adolescence and young adulthood for males in all country groupings. Rates of alcohol-use disorders are higher in males in all country strata, although in high-income countries, the gender gap is less. In high-income countries, the highest rates of substance use generally coincide with the peak in first parenthood for both sexes.

There is consistent and clear evidence that persisting maternal tobacco, alcohol, cannabis and other illicit drug use during pregnancy adversely affects growth and development of the offspring[58]. Maternal smoking has consistently been linked to adverse birth and child outcomes ranging from poor fetal growth, low birth weight, stillbirth, sudden unexpected death in infancy and a broad range of birth defects to later childhood

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

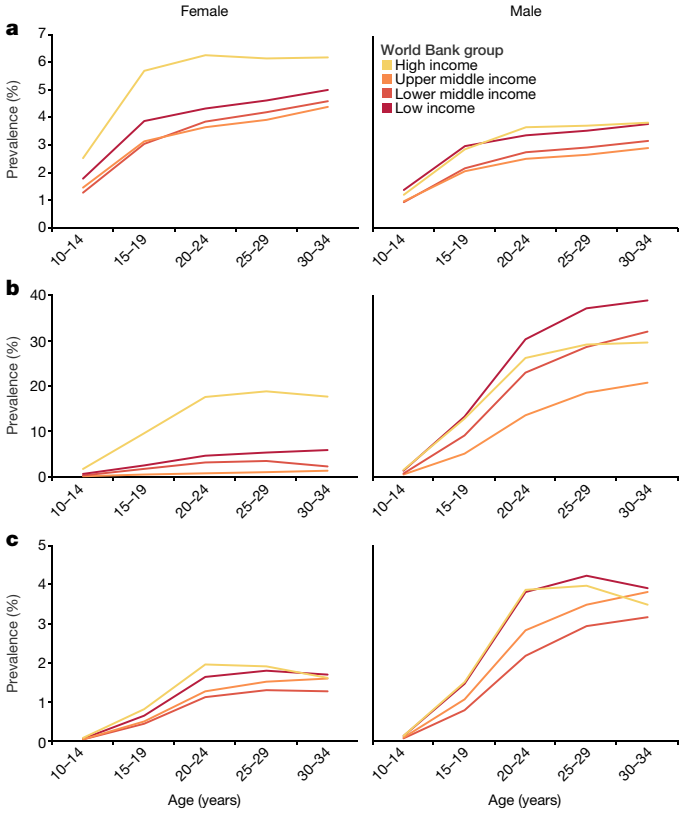


Figure 3 | Mental health and substance use risks across the years of transition to first parenthood for 2016. a, Prevalence of major depression and dysthymia in females and males. **b**, Prevalence of daily tobacco use in females and males. **c**, Prevalence of alcohol-use disorders in females and males.

behavioural problems, obesity and impaired lung function[98,99]. Heavier antenatal drinking (that is, more than three standard drinks per day) is also independently predictive of poor birth outcomes. What the risks of lower levels of consumption are remains debated[100]. Animal models indicate that alterations in fetal growth occur even with brief periconceptional alcohol exposure[101] and even low levels of alcohol consumption during the first trimester affect craniofacial development in humans[102].

Women commonly modify their substance use during later pregnancy with modest benefits for fetal growth and development[103]. Heavy and dependent users, more common in countries with delayed parenthood, are less likely to cease use during pregnancy[104]. However, given that around 40% of pregnancies are unintended and recognition typically occurs around 6–8 weeks of gestation, pre- and periconceptional exposure is likely to be very high in countries in which substance use in adolescents is common[105,106].

Obesity and metabolic disruption. Figure 2b and Extended Data Figure 2 illustrate the rapid increase in obesity during adolescence and young adulthood in females across all country income strata and for males in high-income and upper middle-income countries. The increase with age reflects obesity's strong tendency to persist once established[107,108]. Substantial increases in rates have occurred since 1990 across all country income groups. In high-income countries, prevalence rates of obesity for both females and males in the peak years for first parenthood are around one in five. Although rates are lower elsewhere, one in ten 25–29-year-old women in upper middle-income countries are obese.

Antenatal maternal obesity predicts macrosomic birth[109], later childhood obesity and metabolic disturbance[110,111], poorer cognitive skills and greater risk of behavioural problems during childhood[112]. Antenatal interventions to prevent these consequences appear to have limited benefits. Exercise has the potential to promote greater maternal insulin sensitivity, but its benefits during pregnancy are debated[113,114]. A recent Cochrane review failed to demonstrate that dietary and lifestyle (including exercise) interventions for women with gestational diabetes mellitus restored glycaemic control or prevented adverse effects for women or their babies[113,115]. Although two reports have suggested some modest benefits of antenatal exercise in obese women, its late timing in those with no regular pattern of physical activity before pregnancy seems likely to bring limited benefits[116,117]. By contrast, exercise and diet before pregnancy may bring greater offspring benefits. Exercise initiated before and continued during pregnancy in female rats at risk of gestational diabetes prevents the development of glucose intolerance (gestational diabetes) by increasing β-cell mass and function[118]. Given rapidly rising rates of obesity in nearly all countries, primary prevention that starts in late childhood and continues during adolescence will be essential to avoid wide-ranging adverse intergenerational effects. Sharp drops in physical activity across the second decade, with only around 20% of adolescents globally achieving recommended levels, suggest that promoting physical activity should be a priority[119,120].

Adolescence and the next generation

Growth in the first thousand days predicts later-life health, human capability and resilience to adversity, which is why policies have focused on the promotion of optimal growth and human capital during the first thousand days. Increasingly it has also become a focus for the prevention of non-communicable diseases[121]. However, many of the processes that shape growth during early life originate well before conception or the first antenatal visit. Whether associated with early and high fertility, or with later and lower fertility, growth and development during adolescence provide a foundation for the start to life of the next generation.

The transition through education to employment typically occurs in adolescence, allowing the acquisition of assets that will be essential for being an effective parent, including financial resources and property, and extending to physical, cognitive, social and emotional capabilities (see Fig. 1). However, there are obstacles for many adolescents in achieving a secure foundation for parenthood. Early marriage is particularly noteworthy. Although rates are generally falling, more than half of girls continue to be married before the age of 18 years across many countries in sub-Saharan Africa[122]. Early marriage predicts early parenthood, overall fertility and poor maternal and child nutrition[123–125]. Its effects on the next generation are further mediated through curtailed education and exclusion of women from the formal workforce[126]. The annual benefits in 2030 from ending child marriage from reduced population growth and reductions in early childhood mortality and stunting have recently been estimated at US$664 billion[126]. For proven interventions targeting social norms to delay early marriage, we estimated their benefit–cost ratios to be 8.9 in terms of improving education in the next generation (see Box 1).

Adolescent investments for the next generation will differ in different places. The education of girls is essential in countries with high and early fertility, and early marriage[127]. The education of boys also appears to be important, in that males without education are more likely to marry girls before the legal age and educated fathers have daughters who marry later[125]. Figure 4 illustrates progress in secondary school completion across country income groups in 1990 and 2010. Over 20 years, rates of secondary school completion for girls and boys have increased in all income groups, although much less so in low-income countries, in which only a small minority of girls complete secondary education. Despite gains in lower middle-income countries, rates of secondary school completion for both boys and girls remain under 50%, underlining the importance of continuing investments throughout the Sustainable Development Goal period. Ultimately, success in promoting education will also depend on tackling adolescent undernutrition, social norms favouring early marriage, exposure to hazardous social and physical environments, and adolescents' poor access to health services, including modern contraception.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

BOX 1

Illustration of the benefits of delaying early marriage for the next generation

A recent investment case of adolescents showed that the promotion of education for girls and delaying marriage produced an almost sixfold increase in benefits across the course of life compared to costs[15]. That analysis excluded the intergenerational benefits for their children. Interventions to extend education and to delay marriage will improve a young woman's nutrition and growth as well as allow the acquisition of assets that extend her and her partner's parenting capabilities[141]. Conversely women who marry and parent at a young age are often trapped in a cycle of poverty with poorer life outcomes for their children through the many processes outlined here[126].

One intergenerational benefit of delaying early marriage is greater education of the next generation[142]. This in turn is likely to have substantial health benefits for that next generation[11]. A recent report that was based on studies in 32 sub-Saharan African countries estimated that the children of girls that marry before the age of 18 years have, on average, 1.8 fewer years of schooling[142]. We used models developed for the recent global investment case for adolescence to estimate benefit–cost ratios for delaying marriage based on the effectiveness of a program targeting social norms in young women. This reduced rates of early marriage by 23.4% over four years[143]. We used estimates of the benefit for offspring, using an estimated 12.2% return per year of schooling for lifetime income, for economies in sub-Saharan Africa[144]. The value of an extra 1.8 years schooling for offspring, achieved by an intervention to defer early marriage, produced a benefit–cost ratio of 8.9 (95% uncertainty, 7.2–10.6). This illustrates the substantial intergenerational gains that are likely to arise from policies that emphasize education and delaying marriage in settings of high and early fertility.

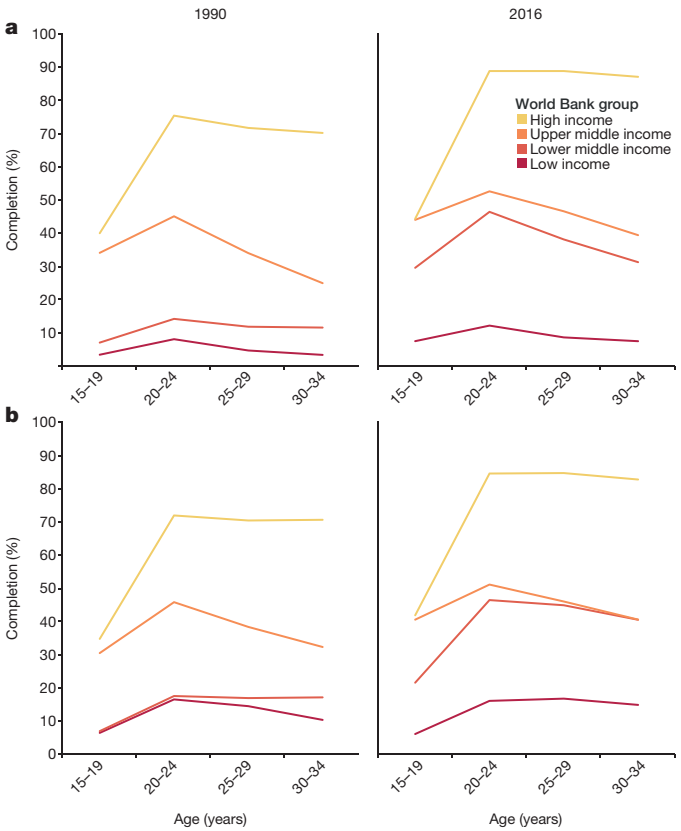


Figure 4 | Secondary school completion (12 years and older) across age and income groups for 1990 and 2010. a, **b**, Secondery school completion for females (**a**) and males (**b**) in 1990 and 2010.

In higher-income countries that have made a transition to low and late fertility, shifting the policy focus to health risks that emerge in adolescence and persist through to parenthood will be essential. Rates of adolescent mental disorders appear to be reaching historic highs and have a strong tendency to persist into the child-bearing years with consequences for maternal and child health, and family wellbeing. Similarly, adolescent and young adult alcohol use and other substance use continue to rise in many countries. Even more marked is the rise in obesity, particularly given that available antenatal interventions to reduce the metabolic consequences for both mothers and offspring are limited[128].

Intergenerational epidemiological studies remain uncommon, and few have addressed risks during sensitive time periods[129]. However, it is becoming clear that different approaches to pregnancy and parenthood will be needed. Preparation for pregnancy and preconception care has been one recent focus[48,75,130]. Previously, the main preconception focus of health services has been on the prevention of unintended pregnancy through provision of contraception[131]. Nevertheless, in countries with high levels of primary care coverage, in which a majority of women make contact with health services at least annually, there would be scope to screen and intervene around intergenerational health risks[132]. The major barriers are the low current demand from future parents and a limited provision of more comprehensive healthcare, particularly for women at the greatest risk[131]. There is little doubt that a reorientation of service systems towards preconception would be of value[133]. However, given continuing high rates of unintended pregnancy[134,135] and the difficulty of modifying chronic health risks, addressing obesity, mental disorders and substance dependence earlier in adolescence will be essential. It requires a broadening of adolescent health beyond a traditional emphasis on sexual and reproductive health, extending the engagement of health service systems with adolescents and creating health-promoting environments in the families, schools, workplaces and communities in which adolescents are growing up[11]. Given the extent to which adolescents shape their own social and nutritional environments, this will require an active engagement with adolescents themselves.

The rapid rise in obesity in middle-income countries, even where undernutrition and food insecurity persist, will require multicomponent approaches with elements addressing healthy diets, physical activity and sedentary behaviour, as well as creating the opportunities for adolescents and their families to make healthy choices[136]. Similarly, high rates of mental disorders are likely to require not only early clinical intervention, but also the targeting of risk factors and the acquisition of protective social and emotional skills. For poor and socially marginalized adolescents in higher-income countries, who often have high and early fertility, responses may need to be similar to those for adolescents growing up in low-income countries[19]. Creating health-promoting environments for adolescents will ultimately require engagement well beyond the health sector, with education, local government, industry, religious leaders, civil society and young people themselves all essential actors.

Girls and young women should undoubtedly remain a priority. However boys and young men should also be brought into focus. They have important roles in parenting that are affected by health problems that commonly emerge before conception[137]. Their values and behaviours affect the capacities of young women to become effective mothers[138]. Increasingly we also understand that their influence on the next generation extends to distinct biological processes that directly affect the early development of the next generation.

Relative to other ages, the current generation of 10–24-year-olds is the largest cohort seen[139]. The potential demographic dividend from a large healthy and educated adolescent cohort entering the workforce has already captured policy attention[25]. We have largely overlooked the

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

fact that this will also be the largest generation to parent, promising an additional dividend in the health, growth and capabilities of the next generation[43]. In a world of competing policy priorities, there is no doubt that providing the resources for healthy adolescent growth, education and emotional development will yield large benefits for current and future generations.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 October 2017; accepted 18 January 2018.

1. Lamy, P. *Now for the Long Term: The Report of the Oxford Martin Commission for Future Generations* (Oxford Univ. Press, 2013).
2. Jamison, D. T. *et al.* Global health 2035: a world converging within a generation. *Lancet* **382,** 1898–1955 (2013).
3. GBD 2015 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388,** 1603–1658 (2016).
4. Schlegel, A & Barry, H. *Adolescence: An Anthropological Enquiry* (Free Press, 1991).
5. Gluckman, P. D. & Hanson, M. A. Evolution, development and timing of puberty. *Trends Endocrinol. Metab.* **17,** 7–12 (2006).
6. Ljung, B.-O., Bergsten-Brucefors, A. & Lindgren, G. The secular trend in physical growth in Sweden. *Ann. Hum. Biol.* **1,** 245–256 (1974).
7. Meng, X., Li, S., Duan, W., Sun, Y. & Jia, C. Secular trend of age at menarche in Chinese adolescents born from 1973 to 2004. *Pediatrics* **140,** e20170085 (2017).
8. Kac, G., Auxiliadora De Santa Cruz Coelho, M. & Velasquez-Melendez, G. Secular trend in age at menarche for women born between 1920 and 1979 in Rio de Janeiro, Brazil. *Ann. Hum. Biol.* **27,** 423–428 (2000).
9. Patton, G. C. & Viner, R. Pubertal transitions in health. *Lancet* **369,** 1130–1139 (2007).
10. Prentice, A. M. *et al.* Critical windows for nutritional interventions against stunting. *Am. J. Clin. Nutr.* **97,** 911–918 (2013).
11. Patton, G. C. *et al.* Our future: a *Lancet* commission on adolescent health and wellbeing. *Lancet* **387,** 2423–2478 (2016).
12. Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D. & Patton, G. C. The age of adolescence. *Lancet Child Adolesc. Health* https://doi.org/10.1016/S2352-4642(18)30022-1 (2018).
13. Viner, R. M. *et al.* Adolescence and the social determinants of health. *Lancet* **379,** 1641–1652 (2012).
14. Mokdad, A. H. *et al.* Global burden of diseases, injuries, and risk factors for young people's health during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **387,** 2383–2401 (2016).
15. Sheehan, P. *et al.* Building the foundations for sustainable development: a case for global investment in the capabilities of adolescents. *Lancet* **390,** 1792–1806 (2017).
16. Pitkin, R. M. Folate and neural tube defects. *Am. J. Clin. Nutr.* **85,** 285S–288S (2007).
17. Peckham, C. & Gibb, D. Mother-to-child transmission of the human immunodeficiency virus. *N. Engl. J. Med.* **333,** 298–303 (1995).
18. Lee, M.-J., Hallmark, R. J., Frenkel, L. M. & Del Priore, G. Maternal syphilis and vertical perinatal transmission of human immunodeficiency virus type-1 infection. *Int. J. Gynaecol. Obstet.* **63,** 247–252 (1998).
19. Aizer, A. & Currie, J. The intergenerational transmission of inequality: maternal disadvantage and health at birth. *Science* **344,** 856–861 (2014).
20. Thompson, O. Genetic mechanisms in the intergenerational transmission of health. *J. Health Econ.* **35,** 132–146 (2014).
21. Godfrey, K. M., Gluckman, P. D. & Hanson, M. A. Developmental origins of metabolic disease: life course and intergenerational perspectives. *Trends Endocrinol. Metab.* **21,** 199–205 (2010).
22. Mollborn, S. Teenage mothers today: what we know and how it matters. *Child Dev. Perspect.* **11,** 63–69 (2017).
23. Rostgaard, T. & Møberg, R. J. in *Fatherhood in the Nordic Welfare states: Comparing Care Policies and Practice* (eds Eydal, G. B. & Rostgaard, T.) Ch. 2, 23–52 (Policy Press, 2014).
24. Tanturri, M. L. in *Contemporary Issues in Family Studies* (eds Abela, A. & Walker, J.) 136–150 (John Wiley & Sons, 2013).
25. World Bank. *World Development Report 2007: Development and the Next Generation* (World Bank, 2007).
26. Hoffman, S. D. & Maynard, R. A. *Kids having Kids: Economic Costs & Social Consequences of Teen Pregnancy* (The Urban Insitute, 2008).
27. Ganchimeg, T. *et al.* Pregnancy and childbirth outcomes among adolescent mothers: a World Health Organization multicountry study. *Br. J. Obstet. Gynaecol.* **121,** 40–48 (2014).
28. Fall, C. H. *et al.* Association between maternal age at childbirth and child and adult outcomes in the offspring: a prospective study in five low-income and middle-income countries (COHORTS collaboration). *Lancet Glob. Health* **3,** e366–e377 (2015).
29. Das, J. K. *et al.* Nutrition in adolescents: physiology, metabolism, and nutritional needs. *Ann. NY Acad. Sci.* **1393,** 21–33 (2017).
30. Akresh, R., Bhalotra, S., Leone, M. & Osili, U. O. War and stature: growing up during the Nigerian Civil War. *Am. Econ. Rev.* **102,** 273–277 (2012).
31. Li, Q. & An, L. Intergenerational health consequences of the 1959–1961 Great Famine on children in rural China. *Econ. Hum. Biol.* **18,** 27–40 (2015).
32. Holmqvist, G. & Pereira. A. Famines and stunting: Are adolescents the hardest hit? https://blogs.unicef.org/evidence-for-action/famines-and-stunting-are-adolescents-the-hardest-hit/ (Unicef, 2017).
33. King, J. C. The risk of maternal nutritional depletion and poor outcomes increases in early or closely spaced pregnancies. *J. Nutr.* **133,** 1732S–1736S (2003).
34. Wu, L. *et al.* Paternal psychological stress reprograms hepatic gluconeogenesis in offspring. *Cell Metab.* **23,** 735–743 (2016).
35. Black, R. E. *et al.* Maternal and child undernutrition: global and regional exposures and health consequences. *Lancet* **371,** 243–260 (2008).
36. O'Dowd, R., Kent, J. C., Moseley, J. M. & Wlodek, M. E. Effects of uteroplacental insufficiency and reducing litter size on maternal mammary function and postnatal offspring growth. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **294,** R539–R548 (2008).
37. Tharner, A. *et al.* Breastfeeding and its relation to maternal sensitivity and infant attachment. *J. Dev. Behav. Pediatr.* **33,** 396–404 (2012).
38. Bhalotra, S. & Rawlings, S. Gradients of the intergenerational transmission of health in developing countries. *Rev. Econ. Stat.* **95,** 660–672 (2013).
39. Ahmed, F., Khan, M. R., Banu, C. P., Qazi, M. R. & Akhtaruzzaman, M. The coexistence of other micronutrient deficiencies in anaemic adolescent schoolgirls in rural Bangladesh. *Eur. J. Clin. Nutr.* **62,** 365–372 (2008).
40. Malabarey, O. T., Balayla, J., Klam, S. L., Shrim, A. & Abenhaim, H. A. Pregnancies in young adolescent mothers: a population-based study on 37 million births. *J. Pediatr. Adolesc. Gynecol.* **25,** 98–102 (2012).
41. Mortimer, J. T. Adolescent Mothers in Later Life. Frank F. Furstenberg, Jr., J. Brooks-Gunn, and S. Philip Morgan. Cambridge University Press, New York, 1987. xiv, 204 pp., illus. $27.95. Human Development in Cultural and Historical Contexts. *Science* **240,** 1060–1061 (1988).
42. McLanahan, S., Tach, L. & Schneider, D. The causal effects of father absence. *Annu. Rev. Sociol.* **39,** 399–427 (2013).
43. Miller, A. R. Motherhood delay and the human capital of the next generation. *Am. Econ. Rev.* **99,** 154–158 (2008).
44. Branum, A. M. & Ahrens, K. A. Trends in timing of pregnancy awareness among US women. *Matern. Child Health J.* **21,** 715–726 (2017).
45. Ayoola, A. B., Nettleman, M. D. & Stommel, M. Time from pregnancy recognition to prenatal care and associated newborn outcomes. *J. Obstet. Gynecol. Neonatal Nurs.* **39,** 550–556 (2010).
46. Owolabi, O. O. *et al.* Comparing the use and content of antenatal care in adolescent and older first-time mothers in 13 countries of west Africa: a cross-sectional analysis of Demographic and Health Surveys. *Lancet Child Adolesc. Health* **1,** 203–212 (2017).
47. Victora, C. G., de Onis, M., Hallal, P. C., Blössner, M. & Shrimpton, R. Worldwide timing of growth faltering: revisiting implications for interventions. *Pediatrics* **125,** e473–e480 (2010).
48. Hanson, M. A., Godfrey, K., Poston, L., Bustreo, F. & Stephenson, J. The Health of the 51%: Women Annual Report of the Chief Medical Officer (2014) *Preconception health* https://www.gov.uk/government/publications/chief-medical-officer-annual-report-2014-womens-health (Department of Health and Social Care, 2015).
49. Soubry, A. Epigenetic inheritance and evolution: a paternal perspective on dietary influences. *Prog. Biophys. Mol. Biol.* **118,** 79–85 (2015).
50. Lane, M., Robker, R. L. & Robertson, S. A. Parenting from before conception. *Science* **345,** 756–760 (2014).
51. Dacheux, J. L. & Dacheux, F. New insights into epididymal function in relation to sperm maturation. *Reproduction* **147,** R27–R42 (2013).
52. Johnson, G. D., Mackie, P., Jodar, M., Moskovtsev, S. & Krawetz, S. A. Chromatin and extracellular vesicle associated sperm RNAs. *Nucleic Acids Res.* **43,** 6847–6859 (2015).
53. Eaton, S. A. *et al.* Roll over Weismann: extracellular vesicles in the transgenerational transmission of environmental effects. *Epigenomics* **7,** 1165–1171 (2015).
54. Chen, Q. *et al.* Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* **351,** 397–400 (2016).
55. Rodgers, A. B., Morgan, C. P., Leu, N. A. & Bale, T. L. Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. *Proc. Natl Acad. Sci. USA* **112,** 13699–13704 (2015).
56. Short, A. K. *et al.* Elevated paternal glucocorticoid exposure alters the small noncoding RNA profile in sperm and modifies anxiety and depressive phenotypes in the offspring. *Transl. Psychiatry* **6,** e837 (2016).
57. Short, A. K. *et al.* Exercise alters mouse sperm small noncoding RNAs and induces a transgenerational modification of male offspring conditioned fear and anxiety. *Transl. Psychiatry* **7,** e1114 (2017).
58. Vassoler, F. M., Byrnes, E. M. & Pierce, R. C. The impact of exposure to addictive drugs on future generations: physiological and behavioral effects. *Neuropharmacology* **76,** 269–275 (2014).
59. Laubenthal, J. *et al.* Cigarette smoke-induced transgenerational alterations in genome stability in cord blood of human F1 offspring. *FASEB J.* **26,** 3946–3956 (2012).
60. Lwow, F., Mędraś, M., Słowińska-Lisowska, M., Jóźków, P. & Szmigiero, L. The effect of occasional alcohol drinking on semen quality and sperm morphology among young and healthy Polish men. *J. Mens Health* **13,** e16–e24 (2017).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

61. Zuccolo, L. *et al.* Pre-conception and prenatal alcohol exposure from mothers and fathers drinking and head circumference: results from the Norwegian Mother–Child Study (MoBa). *Sci. Rep.* **7,** 39535 (2016).
62. Huypens, P. *et al.* Epigenetic germline inheritance of diet-induced obesity and insulin resistance. *Nat. Genet.* **48,** 497–499 (2016).
63. McPherson, N. O., Owens, J. A., Fullston, T. & Lane, M. Preconception diet or exercise intervention in obese fathers normalizes sperm microRNA profile and metabolic syndrome in female offspring. *Am. J. Physiol. Endocrinol. Metab.* **308,** E805–E821 (2015).
64. Donkin, I. *et al.* Obesity and bariatric surgery drive epigenetic variation of spermatozoa in humans. *Cell Metab.* **23,** 369–378 (2016).
65. Hosken, D. J. & Hodgson, D. J. Why do sperm carry RNA? Relatedness, conflict, and control. *Trends Ecol. Evol.* **29,** 451–455 (2014).
66. Holman, L. & Price, T. A. Even more functions of sperm RNA: a response to Hosken and Hodgson. *Trends Ecol. Evol.* **29,** 648–649 (2014).
67. Gluckman, P. D., Hanson, M. A. & Spencer, H. G. Predictive adaptive responses and human evolution. *Trends Ecol. Evol.* **20,** 527–533 (2005).
68. Clarke, H. J. & Vieux, K.-F. Epigenetic inheritance through the female germ-line: the known, the unknown, and the possible. *Semin. Cell Dev. Biol.* **43,** 106–116 (2015).
69. Rando, O. J. & Simmons, R. A. I'm eating for two: parental dietary effects on offspring metabolism. *Cell* **161,** 93–105 (2015).
70. Jungheim, E. S. *et al.* Diet-induced obesity model: abnormal oocytes and persistent growth abnormalities in the offspring. *Endocrinology* **151,** 4039–4046 (2010).
71. Ashworth, C. J., Toma, L. M. & Hunter, M. G. Nutritional effects on oocyte and embryo development in mammals: implications for reproductive efficiency and environmental sustainability. *Phil. Trans. R. Soc. B* **364,** 3351–3361 (2009).
72. Saben, J. L. *et al.* Excess maternal fructose consumption increases fetal loss and impairs endometrial decidualization in mice. *Endocrinology* **157,** 956–968 (2016).
73. Stringer, J. M., Barrand, S. & Western, P. Fine-tuning evolution: germ-line epigenetics and inheritance. *Reproduction* **146,** R37–R48 (2013).
74. Fleming, T. P. *et al.* Do little embryos make big decisions? How maternal dietary protein restriction can permanently change an embryo's potential, affecting adult health. *Reprod. Fertil. Dev.* **27,** 684–692 (2015).
75. Hanson, M. A. *et al.* The International Federation of Gynecology and Obstetrics (FIGO) recommendations on adolescent, preconception, and maternal nutrition: "Think Nutrition First". *Int J. Gynaecol. Obstet.* **131,** S213–S253 (2015).
76. Tobi, E. W. *et al.* DNA methylation signatures link prenatal famine exposure to growth and metabolism. *Nat. Commun.* **5,** 5592 (2014).
77. Dominguez-Salas, P. *et al.* Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nat. Commun.* **5,** 3746 (2014).
78. Finer, S. *et al.* Is famine exposure during developmental life in rural Bangladesh associated with a metabolic and epigenetic signature in young adulthood? A historical cohort study. *BMJ Open* **6,** e011768 (2016).
79. Kühnen, P. *et al.* Interindividual variation in DNA methylation at a putative POMC metastable epiallele is associated with obesity. *Cell Metab.* **24,** 502–509 (2016).
80. Ng, M. *et al.* Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384,** 766–781 (2014).
81. Salamonsen, L. A., Evans, J., Nguyen, H. P. & Edgell, T. A. The microenvironment of human implantation: determinant of reproductive success. *Am. J. Reprod. Immunol.* **75,** 218–225 (2016).
82. Greening, D. W., Nguyen, H. P., Elgass, K., Simpson, R. J. & Salamonsen, L. A. Human endometrial exosomes contain hormone-specific cargo modulating trophoblast adhesive capacity: insights into endometrial–embryo interactions. *Biol. Reprod.* **94,** 1–15 (2016).
83. Currie, J. & Moretti, E. *Mother's Education and the Intergenerational Transmission of human Capital: Evidence from College Openings and Longitudinal Data.* Working Paper No. 9360 (National Bureau of Economic Research, 2002).
84. Roth, T. L., Lubin, F. D., Funk, A. J. & Sweatt, J. D. Lasting epigenetic influence of early-life adversity on the *BDNF* gene. *Biol. Psychiatry* **65,** 760–769 (2009).
85. Stein, A. *et al.* Effects of perinatal mental disorders on the fetus and child. *Lancet* **384,** 1800–1819 (2014).
86. Patton, G. C. *et al.* Prediction of perinatal depression from adolescence and before conception (VIHCS): 20-year prospective cohort study. *Lancet* **386,** 875–883 (2015).
87. Pearson, R. M. *et al.* Maternal depression during pregnancy and the postnatal period: risks and possible mechanisms for offspring depression at age 18 years. *JAMA Psychiatry* **70,** 1312–1319 (2013).
88. Parsons, C. E., Young, K. S., Rochat, T. J., Kringelbach, M. L. & Stein, A. Postnatal depression and its effects on child development: a review of evidence from low- and middle-income countries. *Br. Med. Bull.* **101,** 57–79 (2012).
89. Surkan, P. J., Kennedy, C. E., Hurley, K. M. & Black, M. M. Maternal depression and early childhood growth in developing countries: systematic review and meta-analysis. *Bull. World Health Organ.* **89,** 608–615 (2011).
90. Cheong, J. N. *et al.* Sex-specific metabolic outcomes in offspring of female rats born small or exposed to stress during pregnancy. *Endocrinology* **157,** 4104–4120 (2016).
91. Gallo, L. A., Tran, M., Moritz, K. M., Jefferies, A. J. & Wlodek, M. E. Pregnancy in aged rats that were born small: cardiorenal and metabolic adaptations and second-generation fetal growth. *FASEB J.* **26,** 4337–4347 (2012).
92. Murray, L., Hipwell, A., Hooper, R., Stein, A. & Cooper, P. The cognitive development of 5-year-old children of postnatally depressed mothers. *J. Child Psychol. Psychiatry* **37,** 927–935 (1996).
93. Collishaw, S. Annual research review: secular trends in child and adolescent mental health. *J. Child Psychol. Psychiatry* **56,** 370–393 (2015).
94. Patton, G. C. *et al.* The prognosis of common mental disorders in adolescents: a 14-year prospective cohort study. *Lancet* **383,** 1404–1411 (2014).
95. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390,** 1211–1259 (2017).
96. Olsson, C. A. *et al.* Drinking patterns of adolescents who develop alcohol use disorders: results from the Victorian Adolescent Health Cohort Study. *BMJ Open* **6,** e010455 (2016).
97. Michaud, P. A., Berchtold, A., Jeannin, A., Chossis, I. & Suris, J. C. Secular trends in legal and illegal substance use among 16 to 20 year old adolescents in Switzerland. *Swiss Med. Wkly* **136,** 318–326 (2006).
98. Banderali, G. *et al.* Short and long term health effects of parental tobacco smoking during pregnancy and lactation: a descriptive review. *J. Transl. Med.* **13,** 327 (2015).
99. Hackshaw, A., Rodeck, C. & Boniface, S. Maternal smoking in pregnancy and birth defects: a systematic review based on 173 687 malformed cases and 11.7 million controls. *Hum. Reprod. Update* **17,** 589–604 (2011).
100. Patra, J. *et al.* Dose–response relationship between alcohol consumption before and during pregnancy and the risks of low birthweight, preterm birth and small for gestational age (SGA)—a systematic review and meta-analyses. *Br. J. Obstet. Gynaecol.* **118,** 1411–1421 (2011).
101. Gårdebjer, E. M., Cuffe, J. S., Pantaleon, M., Wlodek, M. E. & Moritz, K. M. Periconceptional alcohol consumption causes fetal growth restriction and increases glycogen accumulation in the late gestation rat placenta. *Placenta* **35,** 50–57 (2014).
102. Muggli, E. *et al.* Association between prenatal alcohol exposure and craniofacial shape of children at 12 months of age. *JAMA Pediatr.* **171,** 771–780 (2017).
103. Lumley, J. *et al.* Interventions for promoting smoking cessation during pregnancy. *Cochrane Database Syst. Rev.* **3,** CD001055 (2009).
104. Schneider, S., Huy, C., Schütz, J. & Diehl, K. Smoking cessation during pregnancy: a systematic literature review. *Drug Alcohol Rev.* **29,** 81–90 (2010).
105. Singh, S., Sedgh, G. & Hussain, R. Unintended pregnancy: worldwide levels, trends, and outcomes. *Stud. Fam. Plann.* **41,** 241–250 (2010).
106. Nykjaer, C. *et al.* Maternal alcohol intake prior to and during pregnancy and risk of adverse birth outcomes: evidence from a British cohort. *J. Epidemiol. Community Health* **68,** 542–549 (2014).
107. Patton, G. C. *et al.* Overweight and obesity between adolescence and young adulthood: a 10-year prospective cohort study. *J. Adolesc. Health* **48,** 275–280 (2011).
108. Venn, A. J. *et al.* Overweight and obesity from childhood to adulthood: a follow-up of participants in the 1985 Australian Schools Health and Fitness Survey. *Med. J. Aust.* **186,** 458–460 (2007).
109. Strutz, K. L., Richardson, L. J. & Hussey, J. M. Preconception health trajectories and birth weight in a national prospective cohort. *J. Adolesc. Health* **51,** 629–636 (2012).
110. Wahlqvist, M. L. *et al.* Early-life influences on obesity: from preconception to adolescence. *Ann. NY Acad. Sci.* **1347,** 1–28 (2015).
111. Yu, Z. *et al.* Pre-pregnancy body mass index in relation to infant birth weight and offspring overweight/obesity: a systematic review and meta-analysis. *PLoS ONE* **8,** e61627 (2013).
112. Van Lieshout, R. J., Taylor, V. H. & Boyle, M. H. Pre-pregnancy and pregnancy obesity and neurodevelopmental outcomes in offspring: a systematic review. *Obes. Rev.* **12,** e548–e559 (2011).
113. Agha-Jaffar, R., Oliver, N., Johnston, D. & Robinson, S. Gestational diabetes mellitus: does an effective prevention strategy exist? *Nat. Rev. Endocrinol.* **12,** 533–546 (2016).
114. Hopkins, S. A. & Cutfield, W. S. Exercise in pregnancy: weighing up the long-term impact on the next generation. *Exerc. Sport Sci. Rev.* **39,** 120–127 (2011).
115. Bain, E. *et al.* Diet and exercise interventions for preventing gestational diabetes mellitus. *Cochrane Database Syst. Rev.* **4,** CD010443 (2015).
116. Russo, L. M., Nobles, C., Ertel, K. A., Chasan-Taber, L. & Whitcomb, B. W. Physical activity interventions in pregnancy and risk of gestational diabetes mellitus: a systematic review and meta-analysis. *Obstet. Gynecol.* **125,** 576–582 (2015).
117. Wang, C. *et al.* A randomized clinical trial of exercise during pregnancy to prevent gestational diabetes mellitus and improve pregnancy outcome in overweight and obese pregnant women. *Am. J. Obstet. Gynecol.* **216,** 340–351 (2017).
118. Mahizir, D. *et al.* Exercise before and during pregnancy is more effective in preventing metabolic disease in females born small fed a high fat diet than exercise during pregnancy only. *Clin. Endocrinol.* **86,** 28 (2017).
119. Olds, T. *et al.* How do school-day activity patterns differ with age and gender across adolescence? *J. Adolesc. Health* **44,** 64–72 (2009).
120. Hallal, P. C. *et al.* Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet* **380,** 247–257 (2012).
121. Bundy, D. A. P. *et al.* Investment in child and adolescent health and development: key messages from *Disease Control Priorities*, 3rd Edition. *Lancet* https://doi.org/10.1016/S0140-6736(17)32417-0 (2017).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

122. Koski, A., Clark, S. & Nandi, A. Has child marriage declined in sub-Saharan Africa? An analysis of trends in 31 countries. *Popul. Dev. Rev.* **43,** 7–29 (2017).
123. Adebowale, S. A., Fagbamigbe, F. A., Okareh, T. O. & Lawal, G. O. Survival analysis of timing of first marriage among women of reproductive age in Nigeria: regional differences. *Afr. J. Reprod. Health* **16,** 95–107 (2012).
124. Fagbamigbe, A. F. & Idemudia, E. S. Survival analysis and prognostic factors of timing of first childbirth among women in Nigeria. *BMC Pregnancy Childbirth* **16,** 102 (2016).
125. Hossain, M. G., Mahumud, R. A. & Saw, A. Prevalence of child marriage among Bangladeshi women and trend of change over time. *J. Biosoc. Sci.* **48,** 530–538 (2016).
126. Wodon, Q. *et al. Economic Impacts of Child Marriage: Global Synthesis Report* (World Bank, 2017).
127. Kalamar, A. M., Lee-Rife, S. & Hindin, M. J. Interventions to prevent child marriage among young people in low- and middle-income countries: a systematic review of the published and gray literature. *J. Adolesc. Health* **59,** S16–S21 (2016).
128. Hanson, M. *et al.* Interventions to prevent maternal obesity before conception, during pregnancy, and post partum. *Lancet Diabetes Endocrinol.* **5,** 65–76 (2017).
129. Rando, O. J. Intergenerational transfer of epigenetic information in sperm. Cold Spring Harb. *Perspect. Med.* **6,** a022988 (2016).
130. Mason, E. *et al.* From evidence to action to deliver a healthy start for the next generation. *Lancet* **384,** 455–467 (2014).
131. M'hamdi, H. I., van Voorst, S. F., Pinxten, W., Hilhorst, M. T. & Steegers, E. A. Barriers in the uptake and delivery of preconception care: exploring the views of care providers. *Matern. Child Health J.* **21,** 21–28 (2017).
132. Shannon, G. D., Alberg, C., Nacul, L. & Pashayan, N. Preconception healthcare delivery at a population level: construction of public health models of preconception care. *Matern. Child Health J.* **18,** 1512–1531 (2014).
133. van Voorst, S. F. *et al.* Effectiveness of general preconception care accompanied by a recruitment approach: protocol of a community-based cohort study (the Healthy Pregnancy 4 All study). *BMJ Open* **5,** e006284 (2015).
134. Finer, L. B. & Zolna, M. R. Shifts in intended and unintended pregnancies in the United States, 2001–2008. *Am. J. Public Health* **104,** S43–S48 (2014).
135. Sedgh, G., Singh, S. & Hussain, R. Intended and unintended pregnancies worldwide in 2012 and recent trends. *Stud. Fam. Plann.* **45,** 301–314 (2014).
136. Draper, C. E. *et al.* Application of Intervention Mapping to develop a community-based health promotion pre-pregnancy intervention for adolescent girls in rural South Africa: Project Ntshembo (Hope). *BMC Public Health* **14,** S5 (2014).
137. Wilson, S. & Durbin, C. E. Effects of paternal depression on fathers' parenting behaviors: a meta-analytic review. *Clin. Psychol. Rev.* **30,** 167–180 (2010).
138. Abramsky, T. *et al.* What factors are associated with recent intimate partner violence? Findings from the WHO multi-country study on women's health and domestic violence. *BMC Public Health* **11,** 109 (2011).
139. Gupta, M. D. *The Power of 1.8 Billion: Adolescents, Youth and the transformation of the future* (United Nations Population Fund, 2014).
140. United Nations. *World Population Prospects The 2017 Revision. Key Findings and Advance Tables.* Report No. ESA/P/WP/248 (United Nations, 2017).
141. Güneş, P. M. The role of maternal education in child health: evidence from a compulsory schooling law. *Econ. Educ. Rev.* **47,** 1–16 (2015).
142. Delprato, M., Akyeampong, K. & Dunne, M. Intergenerational education effects of early marriage in sub-Saharan Africa. *World Dev.* **91,** 173–192 (2017).
143. Pande, R., Kurz, K., Walia, S., MacQuarrie, K. & Jain, S. Improving the reproductive health of married and unmarried youth in India: evidence of effectiveness and costs from community-based interventions. Final Report of the Adolescent Reproductive Health Program in India. (ICRW, 2006).
144. Montenegro, C. E. & Patrinos, H. A. *Comparable Estimates of Returns to Schooling around the World*. Policy Research Working Paper WPS7020 (World Bank, 2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Reid for his assistance with the design of Fig. 1. G.C.P., P.S.A. and N.R. are supported by NH and MRC research fellowships. We acknowledge the support of the Victorian Government's Operational Infrastructure Support Program. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the European Commission.

Author Contributions G.C.P. conceptualized the paper in conjunction with other authors. G.C.P. compiled the first draft with written contributions from S.M.S., C.A.O., V.S., R.S., M.E.W. and E.S. G.C.P., P.S.A., V.S., M.S., B.R., K.F. and K.S. contributed to the presented analyses. C.J.L.M., A.H.M. and N.J.K. provided estimates from Global Burden of Disease Study 2016. All authors reviewed first and subsequent versions of the paper and approved the final version.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to G.P. (gcpatton@unimelb.edu.au).

This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

The analysis of risks included the 195 countries represented in the Institute for Health Metrics and Evaluation (IHME) Global Health Data Exchange database. The IHME provide population data for each country by year, sex and age group. We grouped countries according to World Bank income groups on gross national income per capita in US dollars: high income, upper middle income, lower middle income, and low income[145]. Estimates of prevalence are included as the percentage of people affected (for example, daily smokers) within each World Bank stratum. Estimates were generated for males and females; for the age groups (years): 10–14, 15–19, 20–24, 25–29 and 30–34; for the years: 1990, 1995, 2000, 2005, 2010 and 2016. The modelling of the data, including simulation of uncertainty bounds, is described elsewhere[146–148].

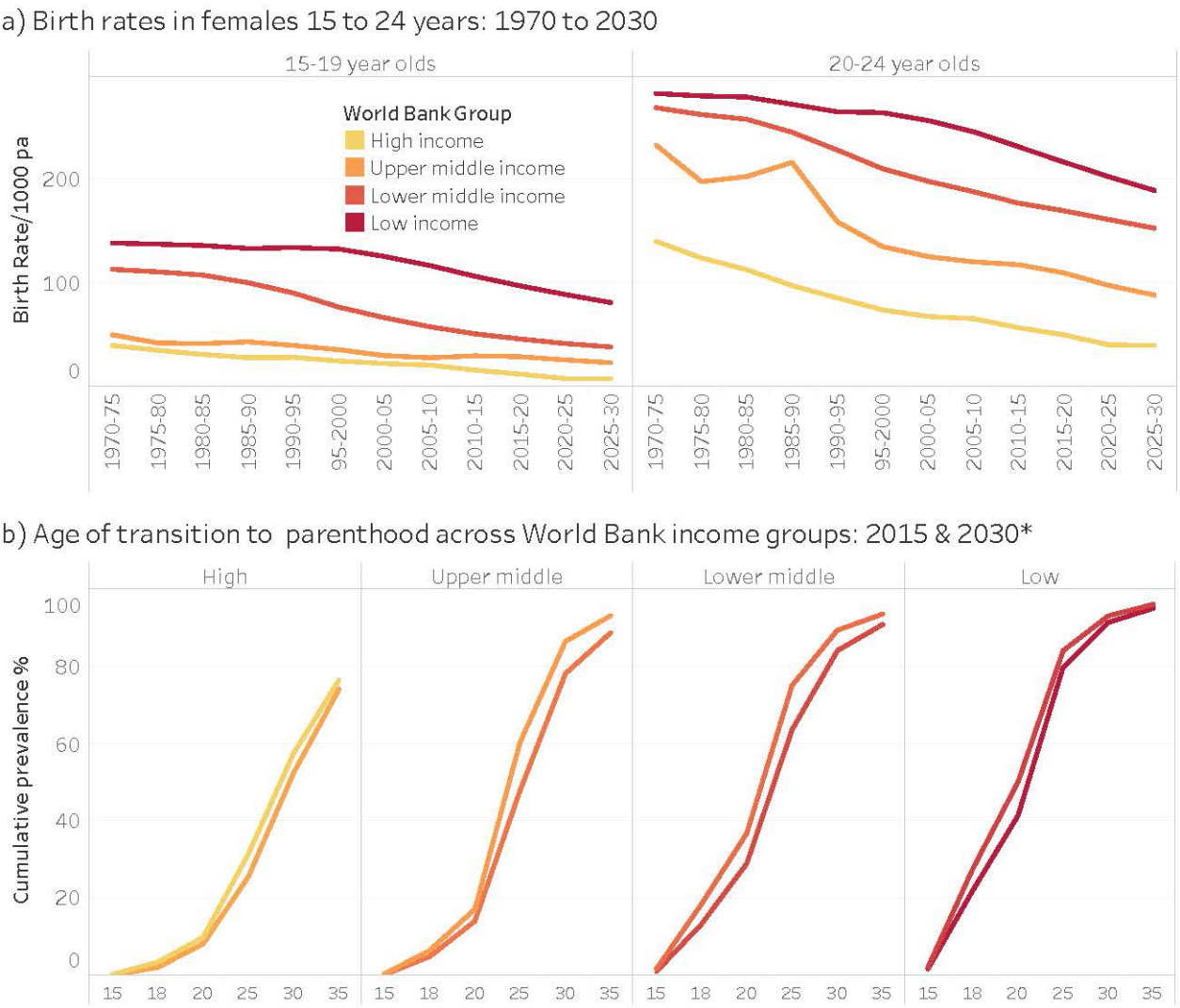
The secondary school completion data were sourced from the database of ref. 149. Data are available for 146 countries. Completion of secondary education was defined as the sum of adolescents who had completed high school together with those currently in tertiary education.

Data in Extended Data Fig. 1 derive from the United Nations Population Division (UNPD) beginning in 1970 and projecting through to 2030 for countries that are stratified by their World Bank income status in 2015. In Fig. 2b, the estimates for 2030 are based on the level of education achieved in 2015 by girls in the respective country strata.

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

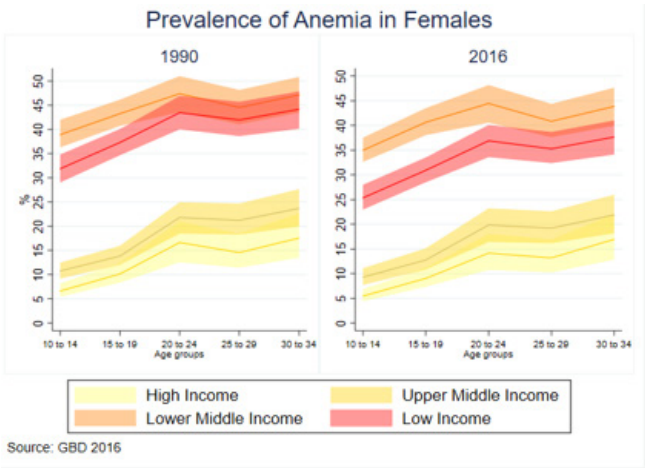
145. World Bank. *How does the World Bank classify countries?* https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries (accessed 6 January 2018).
146. GBD 2016 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390,** 1260–1344 (2017).
147. GBD 2016 SDG Collaborators. Measuring progress and projecting attainment on the basis of past trends of the health-related Sustainable Development Goals in 188 countries: an analysis from the Global Burden of Disease Study 2016. *Lancet* **390,** 1423–1459 (2017).
148. GBD 2015 Obesity Collaborators. Health effects of overweight and obesity in 195 countries over 25 years. *N. Engl. J. Med.* **377,** 13–27 (2017).
149. Barro, R. J. & Lee, J. W. A new data set of educational attainment in the world, 1950–2010. *J. Dev. Econ.* **104,** 184–198 (2013).

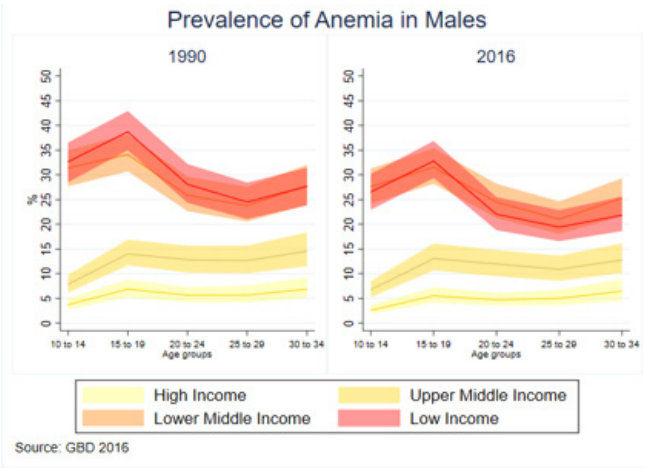
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

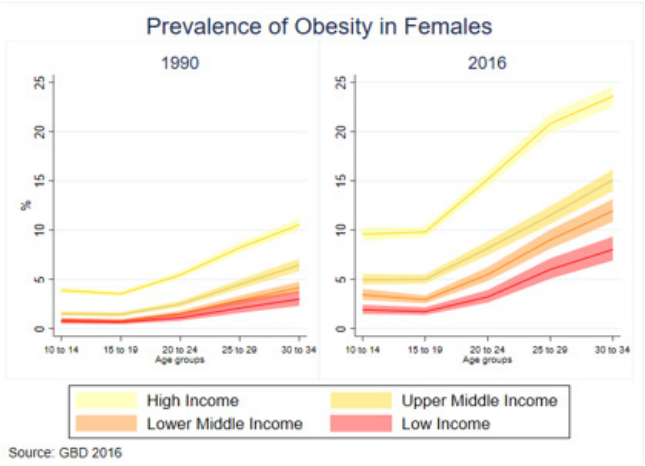


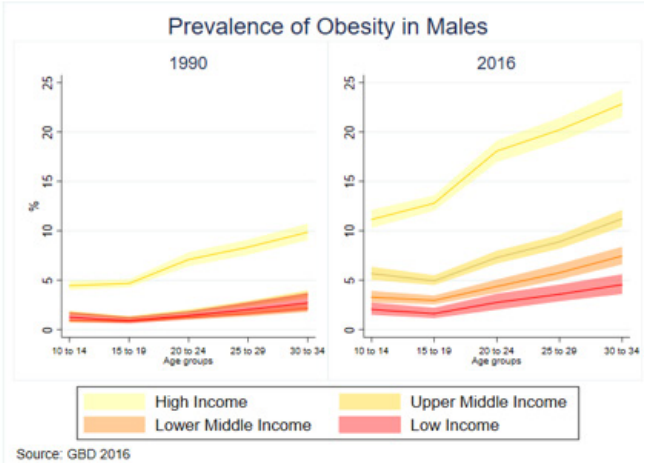
Extended Data Figure 1 | Birth rates and age of transition to parenthood in females across World Bank income groups. a, Trends in birth rates in 15–24 year olds for 1970–2030. **b**, Transition to parenthood in 2015 and 2030. *Projections to 2030 are based on educational attainment of girls in 2015.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



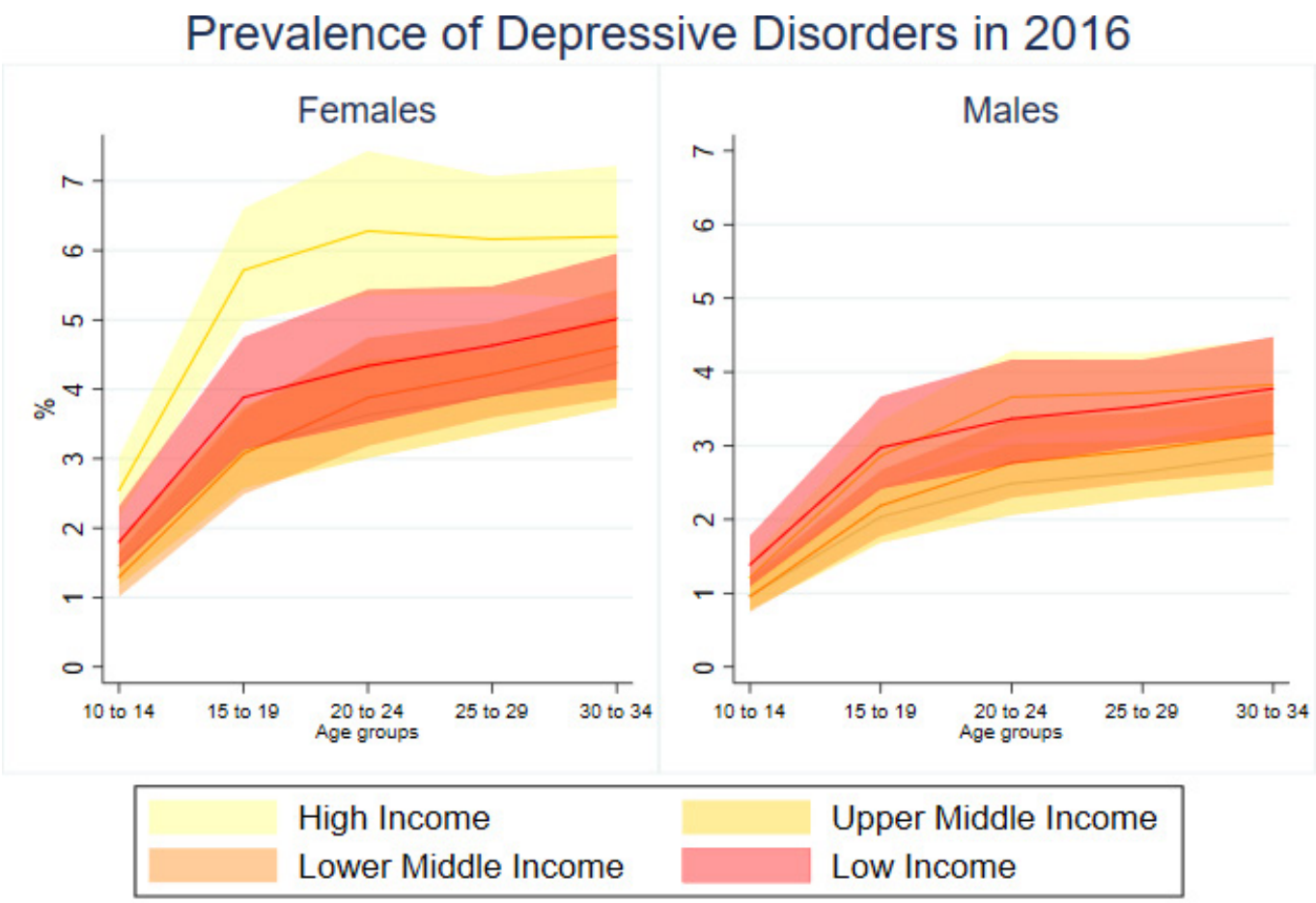






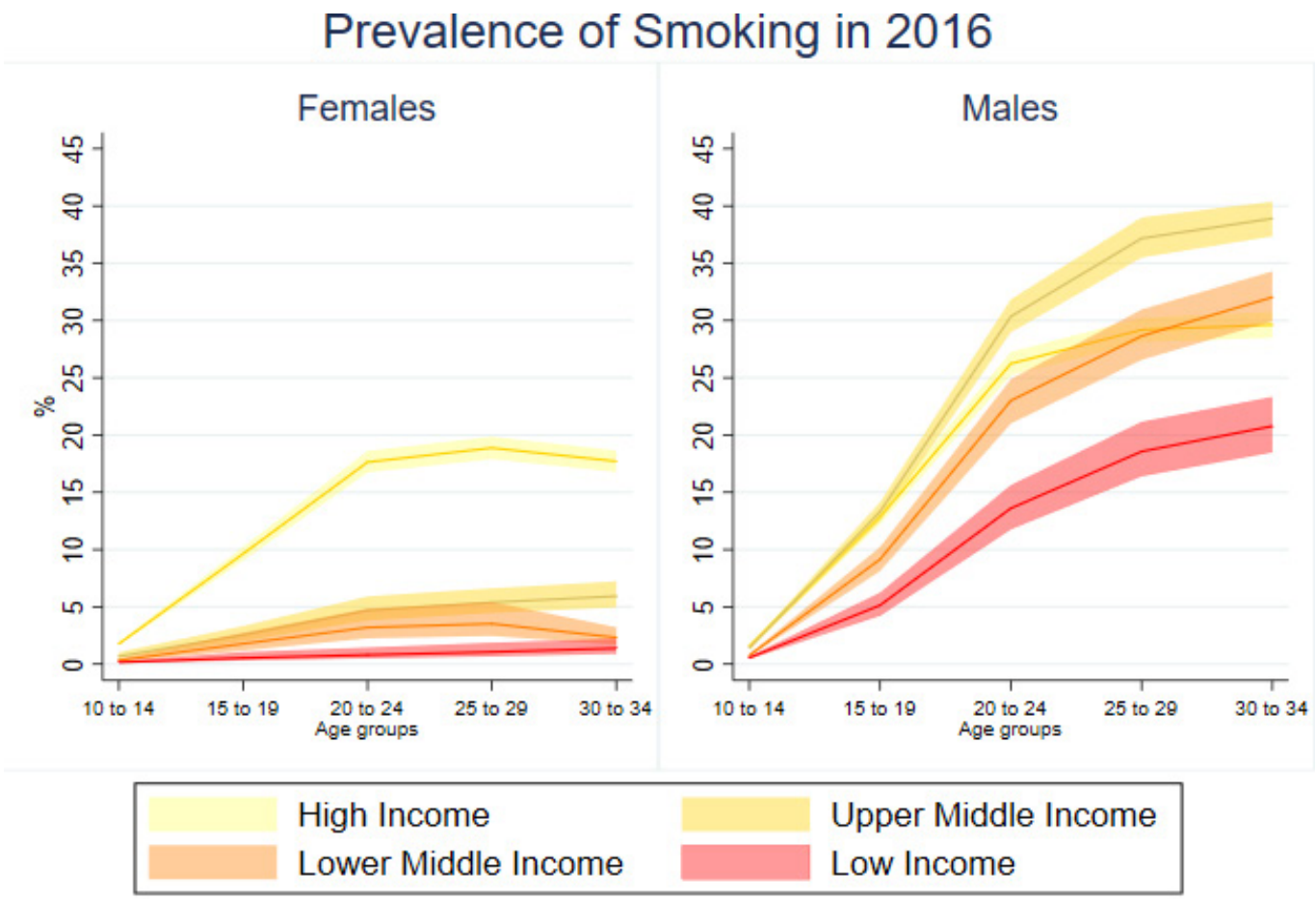
Extended Data Figure 2 | Prevalence of anaemia and obesity in 10–34-year-old females and males in 1990 and 2016 for the four World Bank Country groupings in 2016. Data are mean prevalence (line) and 95% uncertainty bounds (shading). Data were obtained from the Global Health Data Exchange (Global Burden of Disease Study 2016; http://ghdx.healthdata.org/gbd-results-tool).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



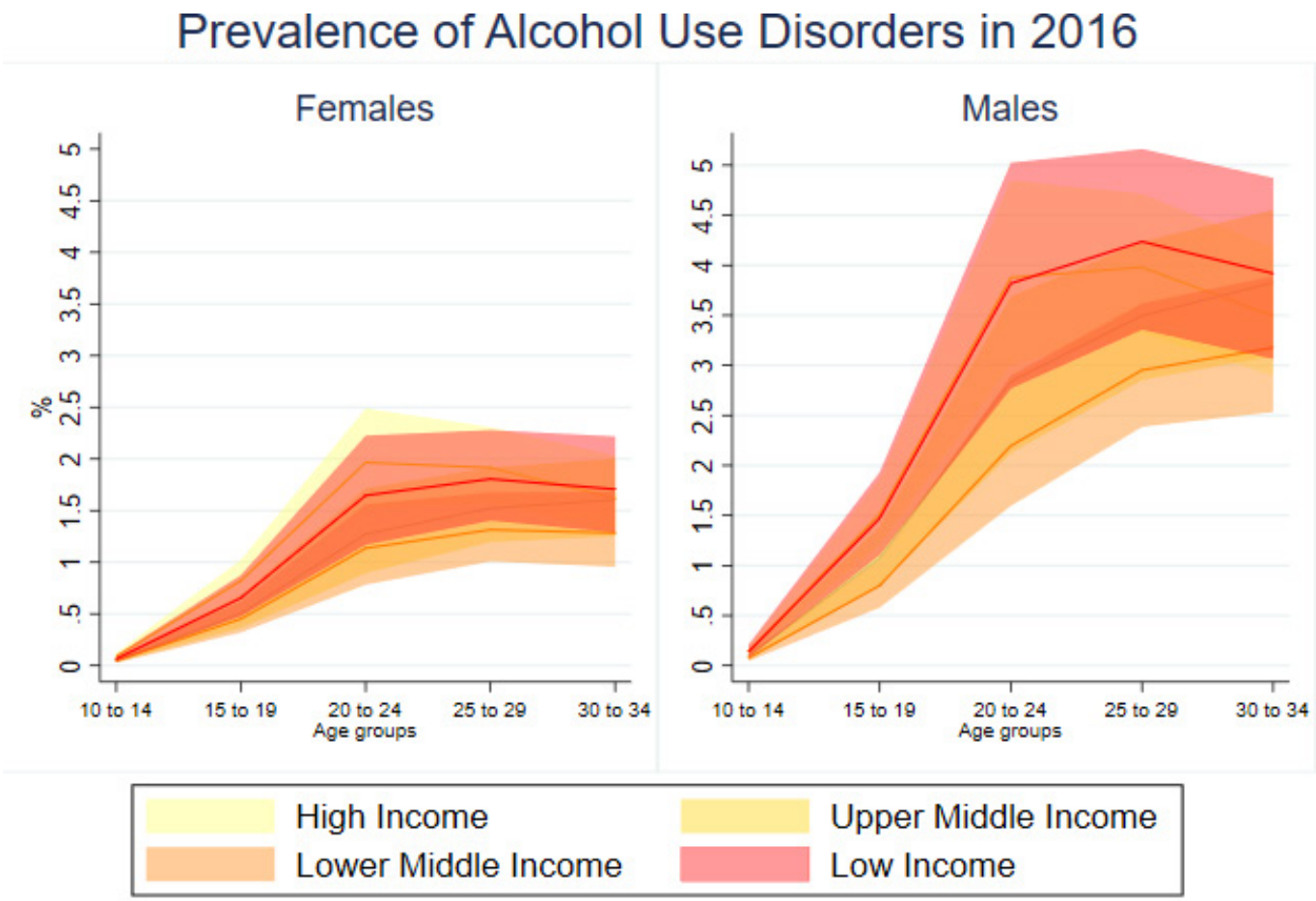
Extended Data Figure 3 | Prevalence of depressive disorders in 10–34-year-old females and males in 2016 for the four World Bank Country groupings. Data are mean prevalence (line) and 95% uncertainty bounds (shading). Data were obtained from the Global Health Data Exchange (Global Burden of Disease Study 2016; http://ghdx.healthdata.org/gbd-results-tool).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 4 | Prevalence of daily smoking in 10–34-year-old females and males in 2016 for the four World Bank Country groupings. Data are mean prevalence (line) and 95% uncertainty bounds (shading). Data were obtained from the Global Health Data Exchange (Global Burden of Disease Study 2016; http://ghdx.healthdata.org/gbd-results-tool).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 5 | Prevalence of alcohol-use disorders in 10–34-year-old females and males in 2016 for the four World Bank Country groupings. Data are mean prevalence (line) and 95% uncertainty bounds (shading). Data were obtained from the Global Health Data Exchange (Global Burden of Disease Study 2016; http://ghdx.healthdata.org/gbd-results-tool).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

MECHANOCHEMISTRY

Molecules pressured to react

Crystals have been made that undergo reactions when compressed. Computational simulations of these processes provide much-needed atomic-level insight into the mechanisms of mechanically induced reactions. SEE LETTER P.505

STUART L. JAMES

Various energy sources can be used to induce chemical reactions — often heat, but also electricity, light and even ultrasound. But there is another option that has been much less explored: mechanical energy from, for example, grinding reactants together or applying pressure to solids. Mechanochemical reactions have been intensively researched in the past few years, but remain poorly understood at the atomic level. On page 505, Yan *et al.*[1] provide some much-needed insight into these processes.

Mechanochemistry has a long history and is currently undergoing a renaissance[2]. The ancient Greek philosopher Theophrastus described how cinnabar (mercury sulfide) can be converted into elemental mercury by grinding it in a copper mortar and pestle with vinegar. In the nineteenth century, Michael Faraday studied mechanochemical reactions[3], and the chemist M. Carey Lea is often credited with demonstrating that mechanically induced reactions can yield different products from thermally induced ones[4]. Polymer chains have also been observed to snap under mechanical stress, and mechanochemical reactions can be expected wherever materials are degraded by mechanical forces[5].

Although mechanochemistry was neglected for many years as a viable method for manufacturing chemicals, in the past decade it has attracted much interest in this regard. This is partly because mechanochemical reactions normally need little or even no solvent, and can therefore save on waste and cost compared with conventional, solvent-based synthetic methods[6]. But despite its growing importance, mechanochemistry remains something of an enigma — exactly how is mechanical energy translated into chemical reactivity?

Yan *et al.* shed some light on this intriguing question. Their work centres on the effect of mechanical force on molecules that consist of carbon-based organic parts connected to an inorganic core of copper and sulfur atoms (Fig. 1). The copper–sulfur core is mechanically 'soft' and susceptible to rupture when a mechanical force is applied; such mechanically sensitive regions of molecules are generally known as mechanophores. By contrast, the organic parts are mechanically 'hard'.

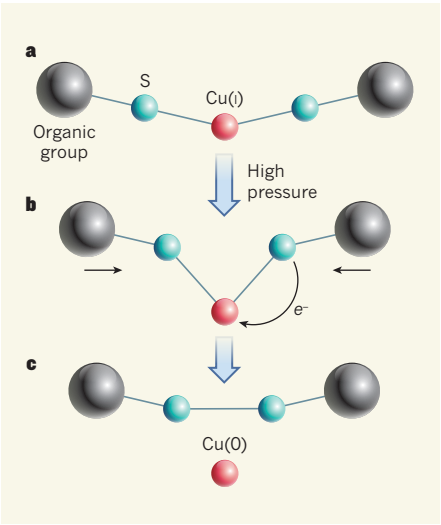


Figure 1 | Reactions induced by compressive force. **a**, Yan *et al.*[1] report molecules that contain incompressible organic groups attached to a 'soft' core consisting of sulfur atoms and copper(I) ions. Only part of the molecule is shown, for simplicity. **b**, When crystals of the molecules are subjected to high pressure (12 gigapascals or more), the organic groups shift towards each other. This alters the bond angles between the atoms in the core, and induces a redox reaction in which electrons (e^-) move from the sulfur atoms to the copper(I) ions. **c**, The reaction products are copper(0) (that is, copper metal) and compounds that contain sulfur–sulfur bonds.

The authors applied pressure of up to 12 gigapascals to crystals of these molecules and observed a reaction in which nanometre-scale particles of metallic copper were produced, as well as organic by-products that contain sulfur–sulfur bonds. In this mechanochemical reaction, electrons are transferred from sulfur to copper, chemically reducing the copper from copper(I) to metallic copper(0). The outcome of the mechanochemical reaction was different from that of heating, which instead produced copper(I) sulfide (Cu_2S) — that is, the mechanochemical and thermochemical reactions broke and formed different chemical bonds. This brings us back to the question of precisely how the mechanochemical process occurs at the atomic level.

Yan and colleagues used computational modelling to show that the application of pressure to the crystals results in 'non-isotropic' compression of the molecules — the mechanically hard organic groups squash the soft copper–sulfur core, dramatically distorting the bond angles between the copper and sulfur atoms. The modelling also showed that, simultaneously with the bond-angle changes, the copper–sulfur bonds become weaker and electron density transfers from the sulfur to the copper atoms. These effects neatly explain the mechanochemical process that leads to the formation of copper-metal particles.

The authors went on to study a variant of their molecule that also contains hard units attached to a soft core but that does not undergo the mechanochemical reaction. Their modelling suggested that the difference in reactivity between the two molecules is due to subtle differences in their structures and crystal packing — for the mechanochemical reaction to occur, the hard organic groups need to have sufficient wiggle room in the crystal to exert force on the copper–sulfur core.

The work is interesting in several further regards. Although a range of mechanophores is known, they are mainly based on organic chemistry and have been explored in the context of polymers. In particular, the rupture of polymer chains caused by stretching can be controlled so that it occurs at specific mechanophore sites[5]. By contrast, Yan and colleagues' mechanophores are inorganic units that undergo redox processes, which occur in response to compression rather than stretching. Moreover, the authors have successfully applied the mechanophore concept to crystals of small molecules, rather than to long-chain polymers. Compressive mechanochemistry such as this might offer a general way to produce nanoparticles.

However, mechanochemical reactions are extremely diverse, and the mechanism put forward in this work helps to explain only a relatively small subset of them. Nevertheless, atomic-level insight into how mechanochemical reactions can occur is invaluable, and Yan and colleagues' work will no doubt stimulate further efforts to reach a similar level of understanding for other reactions, such as those relevant to the synthesis of commercial materials and pharmaceuticals. The authors'

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

use of modelling is particularly effective, and gets around the practical difficulties of directly observing mechanochemical reactions at atomic resolution using microscopy.

Finally, the work points to a future in which an informed understanding of mechanochemistry will improve the prediction and design of reactions. This would provide a more rational basis for the field and facilitate the application of mechanochemical reactions as a mainstream synthetic method. ■

Stuart L. James *is in the School of Chemistry and Chemical Engineering, Queen's University Belfast, Belfast BT9 5GE, UK.*
e-mail: s.james@qub.ac.uk

1. Yan, H. *et al. Nature* **554,** 505–510 (2018).
2. Takacs, L. *Chem. Soc. Rev.* **42,** 7649–7659 (2013).
3. Faraday, M. *Q. J. Sci. Lit. Arts* **8,** 374–376 (1820).
4. Lea, M. C. *Am. J. Sci.* **43,** 527–531 (1892).
5. May, P. A. & Moore, J. S. *Chem. Soc. Rev.* **42,** 7497–7506 (2013).
6. James, S. L. *et al. Chem. Soc. Rev.* **41,** 413–447 (2012).

STRUCTURAL BIOLOGY

Force-activated ion channels in close-up

Piezo proteins allow cells to sense forces by letting ions pass through the cell membrane in response to mechanical stimuli. Three structures of a Piezo protein shed light on how this crucial process works. SEE ARTICLES P.481 & P.487

YUH NUNG JAN & LILY YEH JAN

All cells can sense mechanical force, but it is not clear exactly how. It is known, however, that two closely related members of the mammalian Piezo family of membrane proteins, Piezo1 and Piezo2, have key roles in many physiological and developmental processes that involve mechanosensitivity, including touch, breathing and vascular development[1]. Two papers in *Nature*, by Saotome *et al.*[2] (page 481) and Zhao *et al.*[3] (page 487), and a third published in *eLife* by Guo and MacKinnon[4], now report structures of Piezo1 from the mouse (mPiezo1), obtained using a technique known as cryogenic electron microscopy (cryo-EM). The structures provide insight into how Piezo proteins might sense and respond to mechanical force.

It has been difficult to identify mechanically activated channels — proteins that enable a cell to sense force by allowing ions to pass through the cell membrane in response to mechanical stimuli. A major advance in this field was the discovery of the mechanosensitive Piezo channels[5], which are found in many invertebrates, vertebrates and plants. The mPiezo1 protein is unusually large compared with most proteins, consisting of 2,547 amino-acid residues, and is structurally unrelated to most other proteins.

A previously reported cryo-EM study[6] provided the first insight into the architecture of mPiezo1. It revealed that the channel is a homotrimer (constructed from three identical subunits) consisting of three 'propeller blades' around a central pore. However, the relatively low resolution of the structure allowed only 14 out of the 38 transmembrane helices in each subunit to be resolved, and none of the amino-acid side chains.

The latest cryo-EM structures[2–4] have significantly higher resolution (about 4 ångströms) and reveal many crucial structural features of mPiezo1. The channel is a triskelion, with each arm (or propeller blade) consisting of nine repeated structural units, each containing four transmembrane α-helices. These arms surround a central pore formed by the two transmembrane α-helices (known as TM37 and TM38) nearest to the protein's carboxyl terminus (Fig. 1).

For each arm, the structures of the three repeating units (which consist of TM25 to TM36) nearest the pore were obtained at higher resolution than the next three units (which consist of TM13 to TM24). Each arm bends at the junction between these two trios of repeating units. Viewed from the extracellular side of the channel, the bend angle is about 100°, giving the channel the appearance of a clockwise spiral. When the trimer is viewed from the side, the bend angle is about 140°, so that the channel looks like an inverted dome. Analyses of the amino-acid sequence of mPiezo1 reported in the three new studies suggest that three additional repeat units (which consist of TM1 to TM12) form from the amino-terminal sequences of the protein and comprise the tip of each arm. However, these distal repeat units could not be resolved in the cryo-EM structures, perhaps because they are too flexible.

There are several notable structural features near the centre of the channel. On the extracellular side, there is a large cap on top of the central pore. On the cytoplasmic side, an α-helix acts as a long 'beam' that connects TM28 to the central pore region. This beam might be involved in the mechanism by which force opens and closes the pore (the

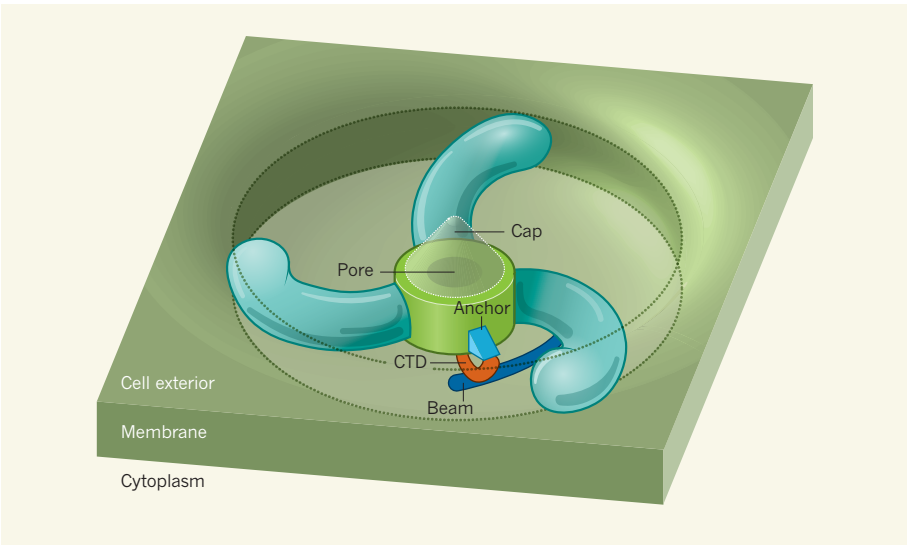


Figure 1 | Cartoon of the mechanosensitive mPiezo1 channel. Three papers[2–4] report the structure of the mouse Piezo1 channel (mPiezo1), a membrane protein that responds to mechanical stimuli by opening its channel, thus allowing external ions to pass through the cell membrane into the cytoplasm. The channel forms from three mPiezo1 proteins, and consists of three bent arms surrounding a central pore, which has an extracellular cap. The cytoplasmic carboxy-terminal domain (CTD) of each mPiezo1 molecule extends from the pore to contact an anchor domain and a 'beam' that connects the arm to the pore. For simplicity, only one CTD, anchor and beam are shown. Guo and MacKinnon[4] propose that the curvature of the arms induces deformation of the cell membrane; the membrane would fill the regions between the arms, but is not shown here for simplicity.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

force-gating mechanism), along with several other domains (dubbed the latch, clasp and anchor) and the carboxy-terminal domain (CTD), which extends from the 'inner helix' formed by the pore-lining TM38. Saotome *et al.* and Zhao *et al.* performed mutagenesis experiments in which they replaced amino acids in these key regions with other amino acids to observe the effects on channel behaviour. These structure-based mutagenesis studies provide insight to the machineries involved in ion permeation through the channel and force gating.

The long central pore of mPiezo1 seems to connect to the aqueous surroundings of the channel through lateral portals that serve as ion-access pathways; additional windows possibly open to the membrane interior. These features are reminiscent of the structures of trimeric acid-sensing ion channels and P2X receptors[7]. Zhao *et al.* report that negatively charged amino-acid residues in the lateral portals influence the ion selectivity and other properties of the pore, supporting the notion that the portals form part of the ion-permeation pathway.

Also notable is Saotome and colleagues' finding that a phenylalanine residue at a pore constriction affects both permeation and the rate at which the channel is inactivated when subjected to a prolonged stimulus. As pointed out by Saotome *et al.* and Zhao *et al.*, inactivation is also modified by disease-causing mutations of multiple charged residues that are involved in the interactions of the anchor with the pore-lining inner helix and the CTD. Taken together with a recent biophysical study[8] that suggests the existence of intricate connections between the ion-permeation pathway and channel inactivation, the structures of mPiezo1 provide hints about the coupling between the pore and the machinery that opens the channel in response to force.

Additional structure-guided studies using mutagenesis and electrophysiology are now needed to further investigate the function of mPiezo1. Other approaches, such as molecular-dynamics simulations, should help to reveal the force-gating mechanisms. Zhao and colleagues' findings, along with those of others[8,9], suggest that multiple parts of mPiezo1 are involved in force gating, including its N-terminal region. A more complete structure that includes the currently unresolved 1,100 amino-acid residues at the N-terminal end of the protein is therefore needed. Moreover, the reported structures seem to show a closed channel. The structure of an open channel could provide valuable clues to the permeation and gating mechanisms.

Guo and MacKinnon observed that the unusual dome shape of the mPiezo1 trimer causes deformation of synthetic membrane vesicles. They suggest that the curvature of the arms induces deformation of the cell membrane that changes with membrane tension and on channel opening. It will be interesting to see whether mPiezo1-induced membrane deformation is detectable *in vivo*, and how it responds to mechanical stimuli.

Apart from mPiezo1, structural information is available for three other families of bona fide mechanically activated channels. The molecules and gating mechanisms all seem to be distinct. The bacterial MscL protein is a homopentamer, and responds to membrane tension by opening a wide pore[10]. The dimeric TRAAK and TREK channels have windows that might allow lipid fatty-acid chains to extend into the pore, and they respond to membrane tension by simultaneously opening the channel and expanding the area of the membrane that is occupied by the channel, although the tension range needed for channel activation differs from that of MscL[11]. The homotetrameric NompC channel found in the fruit fly *Drosophila* contains a bundle of four helices that looks like a coiled spring; this is thought to tether the channel to microtubule structures in the cytoskeleton for force gating[12]. It is to be hoped that some general principles that unite these seemingly disparate structures will emerge as we learn more about each system. ■

Yuh Nung Jan *and* **Lily Yeh Jan** *are at the Howard Hughes Medical Institute, Departments of Physiology and Biochemistry, University of California, San Francisco, San Francisco, California 94158, USA.*
e-mails: yuhnung.jan@ucsf.edu; lily.jan@ucsf.edu

1. Murthy, S. E., Dubin, A. E. & Patapoutian, A. *Nature Rev. Mol. Cell. Biol.* **18,** 771–783 (2017).
2. Saotome, K. *et al. Nature* **554,** 481–486 (2018).
3. Zhao, Q. *et al. Nature* **554,** 487–492 (2018).
4. Guo, Y. R. & MacKinnon, R. *eLife* **6,** e33660 (2017).
5. Coste, B. *et al. Science* **330,** 55–60 (2010).
6. Ge, J. *et al. Nature* **527,** 64–69 (2015).
7. Kellenberger, S. & Grutter, T. *Mol. Biol.* **427,** 54–66 (2015).
8. Moroni, M., Servin-Vences, M. R., Fleischer, R. & Lewin, G. R. Preprint at bioRxiv http://dx.doi.org/10.1101/156489 (2017).
9. Wu, J., Goyal, R. & Grandl, J. *Nature Commun.* **7,** 12939 (2016).
10. Kung, C. *Nature* **436,** 647–654 (2005).
11. Brohawn, S. G. *Ann. N. Y. Acad. Sci.* **1352,** 20–32 (2015).
12. Jin, P. *et al. Nature* **547,** 118–122 (2017).

This article was published online on 12 February 2018.

PANCREATIC DISEASE

An inflammatory transcriptional switch

The mouse pancreas adopts a pre-inflammatory state in response to a chemical injury or the loss of one copy of the gene *Nr5a2*. This state might predispose mice, and possibly humans, to pancreatitis and pancreatic cancer. SEE LETTER P.533

L. CHARLES MURTAUGH & RAYMOND J. MACDONALD

Inflammation can initiate the repair of injured tissue, but, if it persists, can also be a fertile ground for disease and cancer[1]. For instance, chronic inflammation is central to pancreatitis, in which digestive enzymes produced by the pancreas's acinar cells are inappropriately activated, causing tissue digestion and cell death[2]. In mice, transient downregulation of acinar-cell-specific transcription factors and their targets aids recovery from pancreatitis, perhaps by reducing further enzyme production[3]. But irreversible downregulation of these factors is associated with tumours in mice[4–6], and people with pancreatitis have an increased risk of pancreatic cancer[2]. Until now, the causal link between pancreatic inflammation, gene-expression changes and cancer has not been understood. On page 533, Cobo *et al.*[7] describe how repurposing of the pancreatic transcription factor Nr5a2 controls the acinar-cell inflammatory response.

The same research group had previously discovered[8] that loss of one of the two copies (alleles) of the *Nr5a2* gene in mice impairs pancreatic regeneration after mild inflammation, and sensitizes acinar cells to a cancer-causing protein. It seemed likely, therefore, that decreased levels of Nr5a2 diminish the resilience of pancreatic cells.

In the current study, Cobo *et al.* explored this theory by analysing NR5A2-protein levels in samples from humans with a type of cancer called pancreatic ductal adenocarcinoma (PDAC). Consistent with their theory, the researchers found that NR5A2 levels in tumour cells tended to be low in people who had a history of pancreatitis. These patients were also more likely to carry a common single-nucleotide mutation linked to the *NR5A2* gene, associated with increased risk of PDAC, than were people whose tumour cells had high levels of NR5A2. These observations support the idea that low NR5A2 levels increase PDAC risk by raising the

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

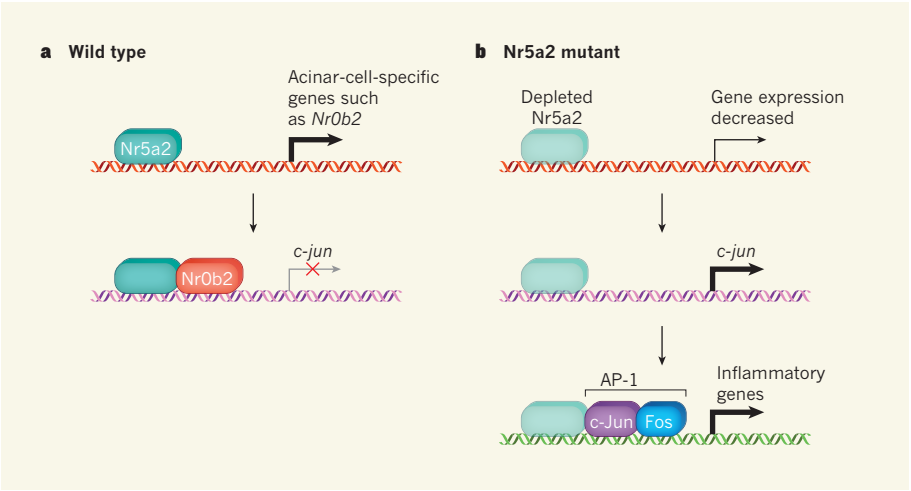


Figure 1 | A transcriptional shift in the pancreas. Cobo *et al.*[7] describe how loss of one of the two copies of the gene *Nr5a2*, which encodes a transcription factor, affects gene expression in the mouse pancreas. **a**, In wild-type animals, the Nr5a2 protein binds to promoter DNA that drives the expression of genes specific to pancreatic acinar cells. These targets include *Nr0b2*, which encodes a transcriptional repressor. Both Nr5a2 and Nr0b2 bind to the promoter of the gene *c-jun*, repressing its expression. **b**, When one copy of *Nr5a2* is deleted, there is less Nr5a2, and so its target genes, including *Nr0b2*, are expressed at lower levels. Under these conditions, Nr0b2 does not associate with Nr5a2, leading to derepression of *c-jun*. The c-Jun protein binds another protein, Fos, to form the transcription factor AP-1. Together, Nr5a2 and AP-1 promote the expression of genes involved in pancreatic inflammation.

propensity for pancreatic inflammation.

Next, the group investigated mice genetically engineered to lack one *Nr5a2* allele. They showed that this mutation triggers a cascade of regulatory events that leads to the upregulation of inflammation-promoting genes. This occurs without infiltration of inflammatory white blood cells into the pancreas, indicating that reduced Nr5a2 expression produces a 'primed' pre-inflammatory state that is a precursor to inflammation. The researchers found that these gene-expression changes are very like those seen shortly after treating mice with caerulein, a chemical inducer of inflammatory pancreatitis in mice.

In a normal pancreas, Nr5a2 binds to promoter DNA sequences that drive the expression of acinar-cell-specific genes to promote cell differentiation, and Cobo and colleagues found that the protein was absent at promoters of inflammatory genes. Therefore, a simple explanation for the pre-inflammatory characteristics observed in Nr5a2-mutant mice might be decreased Nr5a2 binding to the promoters of one or more anti-inflammatory target genes, and so reduced expression of these genes. However, the reality is much more interesting — the authors observed recruitment of Nr5a2 to the promoters of inflammatory genes in mutants.

Many of these inflammatory genes are targets of AP-1, a transcription factor consisting of two proteins: one from each of the Jun and Fos families. Cobo *et al.* found that Nr5a2 interacts physically with one Jun protein, c-Jun, and co-occupies AP-1 target sites on DNA. They then demonstrated the importance of this interaction by showing that pancreas-specific deletion of c-Jun restores Nr5a2 binding to its typical target genes. Furthermore, they showed that Nr5a2-mutant mice are hypersensitive to caerulein-induced pancreatitis, and that this hypersensitivity can be eliminated by *c-jun* deletion. Next, the authors found that the switch of Nr5a2 to AP-1 binding sites also occurs in wild-type mice during early pancreatitis — this, too, depends on c-Jun.

How might reduced expression of a transcription factor cause its redistribution across target genes? One target of Nr5a2 in uninjured tissue is the gene *Nr0b2* (ref. 9), which encodes a transcriptional co-repressor. Cobo *et al.* demonstrate that, in normal conditions, both Nr5a2 and Nr0b2 bind directly to the *c-jun* promoter, ensuring that the gene is only weakly expressed (by contrast, no Nr0b2 binding occurs at Nr5a2's highly expressed targets). When Nr5a2 levels are low, as in mutant mice, Nr0b2 levels decrease rapidly, such that it no longer associates with Nr5a2 at the *c-jun* promoter. The loss of Nr0b2 repressor activity results in increased c-Jun expression, and so higher levels of AP-1. AP-1 then recruits Nr5a2, even at its diminished levels, to different, inflammatory target genes (Fig. 1).

The authors found that, although AP-1 is present at modest levels in the wild-type, uninjured pancreas, it does not bind Nr5a2. Perhaps Nr5a2 binding to Nr0b2 blocks its interaction with c-Jun, and the loss of Nr0b2 during inflammation exposes a high-affinity site for c-Jun, enabling recruitment of Nr5a2 by AP-1. Biochemically focused studies will be required to unravel the complex and



50 Years Ago

Hovermarine, a small firm based in Southampton, has now produced its first hovercraft … The design is particularly interesting; the hovercraft has submerged sidewalls, so that in motion it looks much like a conventional boat … The makers suggest that this sort of craft will be best suited to use in estuaries or rivers, where very large seas are not often experienced. At a speed of 28 knots, the craft will comfortably accommodate seas of up to 3 ft. in height; in more severe conditions, the company says, the maximum speed attainable will depend on the state of the sea and the amount of discomfort which the passengers are prepared to accept.
From *Nature* 24 February 1968

100 Years Ago

Mr. T. J. Westropp … has republished from the Proceedings of the Royal Irish Academy … a paper entitled "The Ancient Sanctuaries of Knockainey and Clogher, Co. Limerick". Here a cairn commemorates the cult of the goddess Aine, of the god-race of the Tuatha de Danann. She was a water spirit, and has been seen, half-raised out of the water, combing her hair. She was a beautiful and gracious divinity … and is crowned with meadowsweet (Spiraea), to which she gave its perfume. She is a powerful tutelary spirit, protector of the sick, and connected with the moon, her hill being sickle-shaped, and men … used to look for the moon — whether risen or not — lest they should be unable to find their way back. They used to visit her shrine on St. John's Eve, carrying wisps of lighted straw, in order to bring good luck to crops and herds … Her son, the magic Earl of Desmond, is still seen riding over the ripples of Loch Gur until his horse's golden shoes are worn out.
From *Nature* 21 February 1918

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

changeable network of Nr5a2 interactions.

Other questions are raised by this work, too. For instance, how does caerulein treatment modulate Nr5a2 function? Cobo *et al.* find that Nr5a2 is redistributed to inflammatory-gene targets within 30 minutes of a single caerulein dose, and that expression of these targets is fully induced within one hour. This implies a rapid mechanism, possibly not dependent on increased c-Jun expression. In addition, this work does not prove that Nr5a2 is required to activate the inflammatory genes that it binds — this will require comparison of mouse pancreases lacking one *Nr5a2* allele with those lacking both (null), in which there is no Nr5a2. It is notable that a previous comparison of *Nr5a2* wild-type and null pancreases did not identify differences in inflammatory-gene expression[9], potentially confirming that residual Nr5a2 is needed to activate inflammatory genes.

Inflammation promotes tumour development in many tissues[1], and Cobo and colleagues' study suggests that altering susceptibility to inflammation may be one way in which common single-nucleotide mutations contribute to the risk of cancer. If future work reveals that human NR5A2 mutations associated with PDAC risk reduce NR5A2 levels in the healthy pancreas, it might be possible to offset the cancer risk using anti-inflammatory drugs. A more targeted approach could take advantage of the potential for NR5A2 to bind and be activated by small organic molecules[10,11]. Such an activating drug might decrease the severity or frequency of pancreatitis in people with insufficient NR5A2, by raising the transcriptional activity of the residual protein. ■

L. Charles Murtaugh *is in the Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112, USA.* **Raymond J. MacDonald** *is in the Hamon Center of Regenerative Science and Medicine and the Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.*
e-mails: murtaugh@genetics.utah.edu; raymond.macdonald@utsouthwestern.edu

1. Grivennikov, S. I., Greten, F. R. & Karin, M. *Cell* **140,** 883–899 (2010).
2. Yadav, D. & Lowenfels, A. B. *Gastroenterology* **144,** 1252–1261 (2013).
3. Murtaugh, L. C. & Keefe, M. D. *Annu. Rev. Physiol.* **77,** 229–249 (2015).
4. Shi, G. *et al. Gastroenterology* **136,** 1368–1378 (2009).
5. Krah, N. M. *et al. eLife* **4,** e07125 (2015).
6. von Figura, G., Morris, J. P., Wright, C. V. E. & Hebrok, M. *Gut* **63,** 656–664 (2014).
7. Cobo, I. *et al. Nature* **554,** 533–537 (2018).
8. Flandez, M. *et al. Gut* **63,** 647–655 (2014).
9. Holmstrom, S. R. *et al. Genes Dev.* **25,** 1674–1679 (2011).
10. Lee, J. M. *et al. Nature* **474,** 506–510 (2011).
11. Sablin, E. J. *et al. J. Struct. Biol.* **192,** 342–348 (2015).

This article was published online on 14 February 2018.

ENGINEERING

Neurons mimicked by electronics

Electronic devices can currently emulate only basic functions of biological neurons. Devices called memtransistors could overcome this limitation and give rise to improvements in artificial-intelligence systems. SEE LETTER P.500

DA LI & XIAOGAN LIANG

The human brain contains billions of neurons that are linked to one another by trillions of tiny contacts called synapses. Designing electronic devices that approach this level of connectivity remains challenging, but, on page 500, Sangwan *et al.*[1] report a breakthrough in this quest. They present a multi-terminal device called a memtransistor, which is made using a single layer of the semiconductor molybdenum disulfide (MoS_2). Such devices could enable the construction of neural-network systems that have a high degree of connectivity and that can execute complex neural functions. This envisaged network technology, if eventually created, would have a great impact on the development of hardware-based artificial intelligence (AI).

For a long time, AI was regarded as science fiction, presented in a variety of forms from humanoid robots to supercomputers. However, the past few years have witnessed rapid progress towards the realization of these fictional machines. For example, in 2016, AlphaGo became the first AI program to defeat a world champion at the game of Go — a victory that greatly increased public interest in AI research (see go.nature.com/2g9kruc).

Neural networks could enable the generation of AI systems that have unprecedented analytical and computational capabilities. However, these networks are typically constructed using electronic components called complementary

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

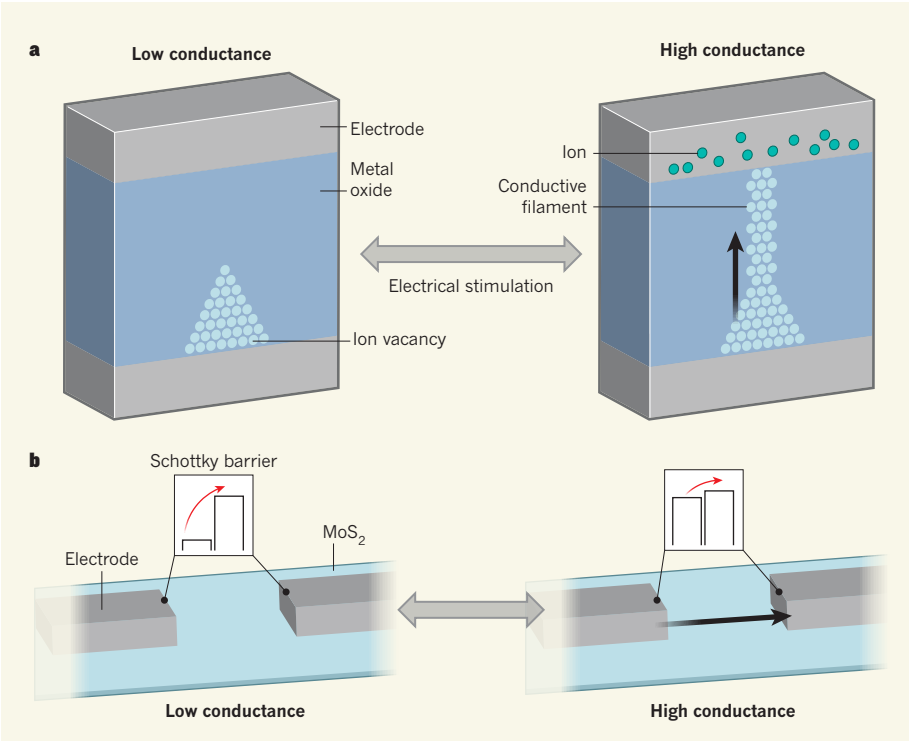


Figure 1 | Neuron-inspired electronic switches. a, Electronic devices known as memristors can emulate basic functions of biological neurons. Conventional memristors consist of a metal-oxide material and two electrodes[3,4]. External electrical stimulation can cause the device to switch between low- and high-conductance states, which are determined by the distribution of internal ions or ion vacancies. In particular, in the high-conductance state, there is a conductive filament of ion vacancies that enables the flow of electric charge[5] (black arrow). **b**, Sangwan *et al.*[1] report a device called a memtransistor that provides a platform potentially capable of carrying out complex neural functions. The device comprises a single layer of molybdenum disulfide (MoS_2) and two or more electrodes. In the low-conductance state, there are high energy barriers called Schottky barriers that charged particles must overcome (red arrow) to travel between electrodes. By contrast, in the high-conductance state, the Schottky barriers are low, which facilitates the flow of electric charge.

metal–oxide–semiconductor (CMOS) devices, which need to be intensively programmed using software. This results in formidable computing complexity and high (megawatt-level) power consumption[2]. These drawbacks severely hinder the practical application of CMOS-based neural-network systems for power-hungry computing tasks, such as multimedia streaming and the control of soft and flexible robots.

To tackle these challenges, researchers are exploring alternative device architectures and principles to those currently used, and different ways to combine (integrate) devices. The aim is to realize energy-efficient AI processes at the hardware level. One approach is to create electronic switches capable of emulating the mechanism that governs signal transmission in the human brain. This mechanism involves changes in the strength of synaptic connections between neurons that correlate with learning and memory[2]. Electronic components known as memristors have become promising candidates for such an application.

Conventional memristors are made from transition-metal oxides[3,4]. In such devices, external electrical stimulation can cause a substantial change in conductance — the device switches between low- and high-conductance states. These states are determined by the distribution of internal ions or ion vacancies, and the switching of the states is attributed to the formation and rupture of a conductive filament that connects two electrodes[5] (Fig. 1a).

A limitation of this filament-based operation scheme is that the memristor can have only two terminals (connections between the metal-oxide material and the electrodes). Although neural networks consisting of two-terminal memristors can execute basic neural functions[5], the human brain contains many more synapses than neurons. This observation suggests that multi-terminal devices based on a non-filament operation principle are needed for carrying out complex neural functions.

"We earnestly hope that, in the near future, small-scale memtransistor networks will be constructed."

The conductance state of Sangwan and colleagues' MoS_2 memtransistor is mainly determined by the heights of energy barriers called Schottky barriers that charged particles must overcome to travel between electrodes (Fig. 1b). External voltages can be used to tune the heights of these barriers and switch between conductance states.

This state-switching principle distinguishes MoS_2 memtransistors from filament-based memristors. In particular, it provides a platform for creating multi-terminal circuits capable of mimicking biological neurons that have many synaptic connections. The authors created a memtransistor that has six electrodes to demonstrate this capability (see Figure 4b of the paper[1]). Because the conductance state between any two electrodes is controlled by voltage-tuned Schottky barriers instead of by filament formation and rupture, the states can be easily tuned by varying the heights of individual barriers. In addition, because the memtransistor uses only a single layer of MoS_2, it is extremely thin — which means that the states can also be tuned using an electrode underneath the device, potentially enabling a larger range of states to be accessible.

Despite the benefits offered by the authors' memtransistor, various challenges must be overcome before such devices can be used to construct working neural networks. For example, industry currently lacks the ability to manufacture large areas of highly uniform single-layer MoS_2. Furthermore, modelling work involving molecular dynamics is needed to fully understand the atomic-level processes responsible for the device's switching characteristics.

Sangwan and colleagues' work is still at the device level, and the authors have not provided a realistic design for a complete neural-network system composed of MoS_2 memtransistors. We earnestly hope that, in the near future, small-scale memtransistor networks will be constructed and used to demonstrate basic neural functions. Such functions could include: pattern classification; a type of learning known as dictionary learning; and feature extraction, in which the number of resources required to describe a large set of data is reduced. Finally, it should be pointed out that biological neurons are arranged in 3D space. This fact could motivate studies that address the challenges associated with the 3D integration of multi-terminal memtransistors. ■

Da Li *and* **Xiaogan Liang** *are in the Mechanical Engineering Department, University of Michigan, Ann Arbor, Michigan 48109, USA.*
e-mail: xiaoganl@umich.edu

1. Sangwan, V. K. *et al. Nature* **554,** 500–504 (2018).
2. Prezioso, M. *et al. Nature* **521,** 61–64 (2015).
3. Williams, R. S. *IEEE Spectrum* **45,** 29–35 (2008).
4. Strukov, D. B., Snider, G. S., Stewart, D. R. & Williams, R. S. *Nature* **453,** 80–83 (2008).
5. Kim, S., Choi, S.-H. & Lu, W. *ACS Nano* **8,** 2369–2376 (2014).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ARTICLE

doi:10.1038/nature25739

A molecular atlas of cell types and zonation in the brain vasculature

Michael Vanlandewijck[1,3]*, Liqun He[2]*, Maarja Andaloussi Mäe[3], Johanna Andrae[3], Koji Ando[3], Francesca Del Gaudio[4], Khayrun Nahar[3], Thibaud Lebouvier[3,5], Bàrbara Laviña[3], Leonor Gouveia[3], Ying Sun[6], Elisabeth Raschperger[1], Markus Räsänen[7], Yvette Zarb[8], Naoki Mochizuki[9,10], Annika Keller[8], Urban Lendahl[4] & Christer Betsholtz[1,3]

Cerebrovascular disease is the third most common cause of death in developed countries, but our understanding of the cells that compose the cerebral vasculature is limited. Here, using vascular single-cell transcriptomics, we provide molecular definitions for the principal types of blood vascular and vessel-associated cells in the adult mouse brain. We uncover the transcriptional basis of the gradual phenotypic change (zonation) along the arteriovenous axis and reveal unexpected cell type differences: a seamless continuum for endothelial cells versus a punctuated continuum for mural cells. We also provide insight into pericyte organotypicity and define a population of perivascular fibroblast-like cells that are present on all vessel types except capillaries. Our work illustrates the power of single-cell transcriptomics to decode the higher organizational principles of a tissue and may provide the initial chapter in a molecular encyclopaedia of the mammalian vasculature.

The profound role of vascular dysfunction in ischaemic stroke and the vascular-specific expression of genes that are mutated in individuals with neurological diseases illustrate the critical importance of vascular health for brain function[1,2]. The brain vasculature harbours a unique specialization—the blood-brain barrier (BBB)—that is necessary for neuronal function[3,4] but represents a hurdle for the pharmacological treatment of brain diseases[5,6]. While specialization of the vascular endothelium is known to be a key feature of the BBB, the contribution of vascular mural cells (pericytes and smooth muscle cells (SMCs)) and vessel-associated cell types such as astrocytes is less clear.

The brain vasculature displays an arteriovenous hierarchy similar to those of other vascular beds. Differences in blood flow, pressure and chemical composition are paralleled along the arteriovenous axis by morphological and functional differences, such as junction structure and permeability, that have been elucidated using electron microscopy[7,8]. Systematic efforts to unravel the molecular basis of vascular arteriovenous specialization have, to our knowledge, not been undertaken, and only a handful of endothelial arteriovenous markers have been identified in various species, organs and developmental stages[9–13]. Mural arteriovenous differentiation involves distinct cellular morphologies ranging from the multiple-layered SMC coats of large arteries to the singular, longitudinally oriented or stellate pericytes of capillaries and venules. However, as for endothelial cells, the molecular underpinnings of these differences are largely unknown. Whereas the contractility of vascular SMCs is crucial for the regulation of blood pressure and flow, the physiological role of pericytes is unclear[14]. Pericyte identification is problematic owing to a lack of specific markers[14], and criteria for pericyte definition remain debated[15].

Considering the vast medical and pharmacological importance of the brain vasculature and the lack of a molecular understanding of its constituent cell types, we set out to transcriptionally profile the principal cell types of the brain vasculature using single-cell RNA sequencing (scRNA-seq).

Cell classes of the brain vasculature

We used a set of transgenic reporter mice to capture all major vascular and vessel-associated cell types from the adult brain (Fig. 1a). Single cells were sorted into 384-well plates (Extended Data Fig. 1a) and profiled using SmartSeq2[16]. About 3,500 single cell transcriptomes were clustered using the BackSPIN clustering algorithm[17] and annotated (Extended Data Fig. 1b–m), leading to the cluster map shown in Fig. 1b. Annotations were guided by the expression of canonical cell class markers, as exemplified in Fig. 1c. Primary bar-plot data and average expression scores for the cell types (Extended Data Fig. 2a) are available gene-by-gene at http://betsholtzlab.org/VascularSingleCells/database.html. This dataset provides molecular definitions for the major vascular and perivascular cell classes of the adult mouse brain. In addition to endothelial and mural cells, we obtained clusters corresponding to astrocytes, oligodendrocyte precursors (*Pdgfra*$^+$), intermediate stage oligodendrocytes (*Opalin*$^+$)[18], perivascular (*Mrc1*$^+$) and interstitial (*Cx3cr1*$^+$) microglia[17,19] and two types of fibroblast-like cells (Extended Data Fig. 2b–d).

Endothelial arteriovenous zonation

Gradual cellular phenotypic changes along an anatomical axis are referred to as cellular 'zonation'[20]. Three of the endothelial clusters (venous (vEC), capillary (capEC) and arterial (aEC)), containing a total of 1,100 endothelial cells, showed a biased distribution of known arteriovenous markers and were re-ordered by SPIN (sorting points into neighbourhoods) into a single one-dimensional range. We found

[1]Karolinska Institutet/AstraZeneca Integrated Cardio Metabolic Centre (KI/AZ ICMC), Blickagången 6, SE-141 57 Huddinge, Sweden. [2]Department of Neurosurgery, Tianjin Medical University General Hospital, Tianjin Neurological Institute, Key Laboratory of Post-Neuroinjury Neuro-Repair and Regeneration in Central Nervous System, Ministry of Education and Tianjin City, Tianjin 300052, China. [3]Department of Immunology, Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Dag Hammarskjölds väg 20, SE-751 85 Uppsala, Sweden. [4]Department of Cell and Molecular Biology, Karolinska Institutet, Von Eulers väg 3, SE-171 77 Stockholm, Sweden. [5]Inserm U1171, University of Lille, CHU, Memory Center, Distalz, F-59000 Lille, France. [6]Department of Bioinformatics, Zhongyuan Union Genetic Technology Co., Ltd., No.45, the 9th East Road, Tianjin Airport Economic Area, Tianjin 300304, China. [7]Wihuri Research Institute and Translational Cancer Biology Program, Biomedicum Helsinki, University of Helsinki, Haartmaninkatu 8, P.O. Box 63, FI-00014 Helsinki, Finland. [8]Division of Neurosurgery, Zürich University Hospital, Zürich University, Zürich, CH-8091, Switzerland. [9]Department of Cell Biology, National Cerebral and Cardiovascular Center Research Institute, Suita, Osaka, Japan. [10]AMED-CREST, National Cerebral and Cardiovascular Center, Suita, Osaka, Japan.
*These authors contributed equally to this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

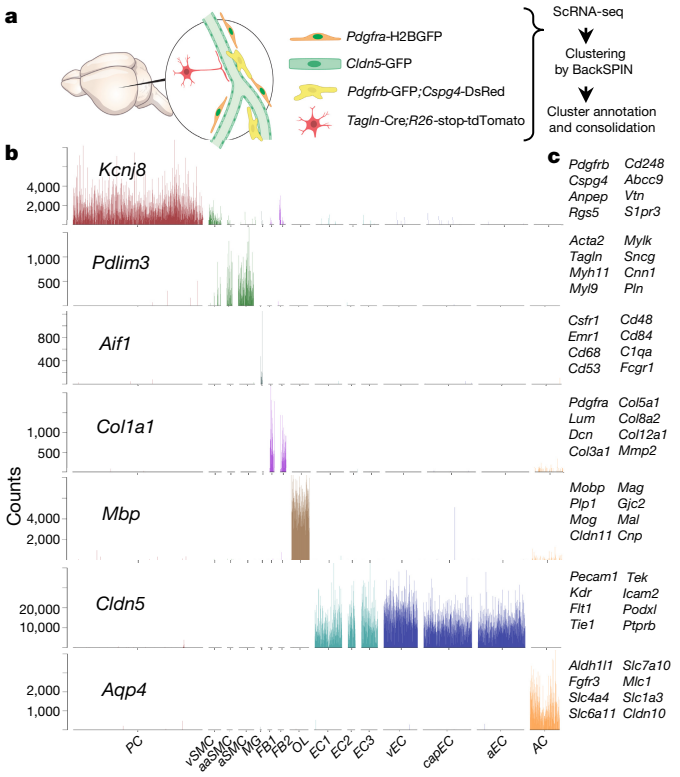


Figure 1 | Procedures and cell clusters. a, Method flowchart. **b**, Bar plots of representative cell type-specific markers. See Extended Data Fig. 1 for cell type abbreviations. **c**, More cell type-specific markers used for cluster annotation. See http://betsholtzlab.org/VascularSingleCells/database.html for high-resolution images and statistics.

that known arterial[9–12] (*Bmx*, *Efnb2*, *Vegfc* and *Sema3g*) and venous[13] (*Nr2f2*) markers peaked at opposite ends of this range (Fig. 2a), with gradual changes in expression that suggested zonation. A total of 1,798 transcripts were significantly differentially expressed across the range (Supplementary Table 1), with the 500 most significant ones (Supplementary Table 2) distributed into six patterns with putative arterial, venous or capillary predominance, or combinations thereof (Fig. 2b). Again, the nested patterns of gradually changing gene expression in the cells across the range were indicative of cellular zonation.

To confirm that the SPIN range represented a cell order matching the arteriovenous axis, we analysed the tissue expression of eight selected markers. *Bmx* was confirmed as a brain arterial marker using *Bmx–lacZ* mice (Extended Data Fig. 3a). *Mfsd2a*, encoding a BBB-specific lipid transporter[21], peaked at the middle of the SPIN range (Fig. 2a), matching the concentration of MFSD2A protein in capillary endothelial cells (Fig. 2c, d and Extended Data Fig. 3b). *Tfrc* (encoding the transferrin receptor (TFRC)) peaked in the middle–left part of the SPIN range (Fig. 2a), matching the TFRC staining of capillaries and veins, but not arteries (Fig. 2c, d and Extended Data Fig. 3c). Similar correlations were found between the expression of *Vwf*, *Vcam1* (artery and vein, but not capillary) and *Slc16a1* (capillary–venous) transcripts and the localization of the corresponding proteins von Willebrand factor (VWF), vascular cell adhesion molecule 1 (VCAM1) and solute carrier 16a1 (SLC16A1) to different vessel types (Fig. 2a, c, d and Extended Data Figs 3d, 4a, b). In addition, RNA *in situ* hybridization (ISH) was used to localize *Gkn3* and *Slc38a5* expression to arteries and veins, respectively, consistent with the SPIN distribution (Fig. 2a and Extended Data Fig. 4c–e). The arteriovenous zonation of these transcripts was also consistent with ISH data deposited in the Allen Brain Atlas[22], which also validated the zonal expression of *Slc6a6* and *Slc7a5* (Extended Data Fig. 5). Collectively, the gradual changes in expression and nested expression patterns of groups of genes across the SPIN range suggest that endothelial arteriovenous zonation represents a seamless continuum of transcriptional states (schematically illustrated in Extended Data Fig. 6a).

As examples of how specific gene or protein classes distribute along the arteriovenous axis, we found zonal and nested distribution of 120 transcription factor transcripts and 145 transmembrane transporter transcripts (Extended Data Fig. 6b, c), providing insight into transcriptional regulation of arteriovenous zonation and how this results in endothelial specialization. We noted a particular difference in distribution: whereas transcription factors were overrepresented at arterial locations, transporters dominated in capillaries and veins, indicating a concentration of BBB-associated trans-endothelial molecular transport functions at the latter locations.

BackSPIN also assigned endothelial cell clusters that fell outside the arteriovenous zone. These clusters (EC1–EC3) (Fig. 1b) showed high expression of ribosomal protein transcripts of the Rpl and Rps classes, suggesting that these cells are (more) active in protein synthesis. Whether EC1–EC3 cells are located in a specific brain region or distributed evenly across brain regions is currently unclear. The presence of arteriovenous markers in all EC1–EC3 clusters suggests that these cells are distributed across all vessel types.

Mural arteriovenous zonation

To study mural cell zonation, we first compared the distribution of known mural cell transcripts across the BackSPIN clusters with the corresponding protein expression *in situ*. *Cnn1*, which was specific to the arterial SMC (aSMC) cluster (Fig. 3a), matched the localization of CNN1 (calponin 1) protein to arteries with diameters larger than 13 μm (Fig. 3b and Extended Data Fig. 6d). *Acta2* and *Tagln*, which were highly expressed in the aSMC and arteriole (aa)SMC clusters (Fig. 3a), matched the expression of α-smooth muscle actin (ACTA2) and smooth muscle protein 22-α (TAGLN) in arteries and arterioles with diameters larger than 8 μm (Extended Data Figs 6d, 7a). Expression of *Acta2* and *Tagln* was weak in the venous (v)SMC cluster and almost absent from the PC (pericyte) cluster, consistent with weak staining for ACTA2 and TAGLN in large veins and undetectable levels in capillaries and venules (Extended Data Fig. 7a). The broad distribution of *Cspg4* and *Pdgfrb* across all mural clusters (Fig. 3a) correlated with the broad expression of DsRed driven by the *Cspg4* promoter (*Cspg4*-DsRed) and GFP driven by the *Pdgfrb* promoter (*Pdgfrb-GFP*) across all vessel types (Extended Data Fig. 7b). Consistent with the mouse data, we found zebrafish *tagln:EGFP* predominantly in brain arteries and arterioles, whereas *pdgfrb*-driven GFP was detected in all vessels (Extended Data Fig. 8a), suggesting that mural cell zonation has been evolutionarily conserved. Consistent with the specific expression of *Abcc9* in the mouse PC and vSMC clusters, the newly established zebrafish line *TgBAC(abcc9:Gal4FF)* showed specific labelling of brain mural cells in capillaries and veins (Fig. 3b and Extended Data Fig. 8b).

Figure 3c illustrates the distribution of the 1,708 most variable mural transcripts (Supplementary Table 3) across 1,385 mural cells arranged by SPIN. Notably, the order of mural cell types according to transcriptional relatedness (PC–vSMC–aaSMC–aSMC) did not match the anatomical organization along the arteriovenous axis. Moreover, hierarchical clustering indicated two predominant gene expression patterns: one cluster of transcripts that was expressed highly in the PC and vSMC clusters but at low levels in the aaSMC and aSMC clusters, and another cluster of transcripts with the opposite pattern. This suggests that there are two distinct subclasses of mural cells (Fig. 3c and Extended Data Fig. 6e): one in which pericytes occur in a continuum with vSMC through gradual loss of PC markers and acquisition of (low levels of) SMC markers, and the other in which aaSMCs are in continuum with aSMCs through the progressive acquisition of aSMC markers. The abrupt transition from aaSMCs to pericytes at the arteriole–capillary boundary is illustrated when the expression of the 1,708 genes are plotted in the correct arteriovenous order (Fig. 3d). High-resolution imaging of ACTA2 and PDGFRB expression in brain tissue shows that

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

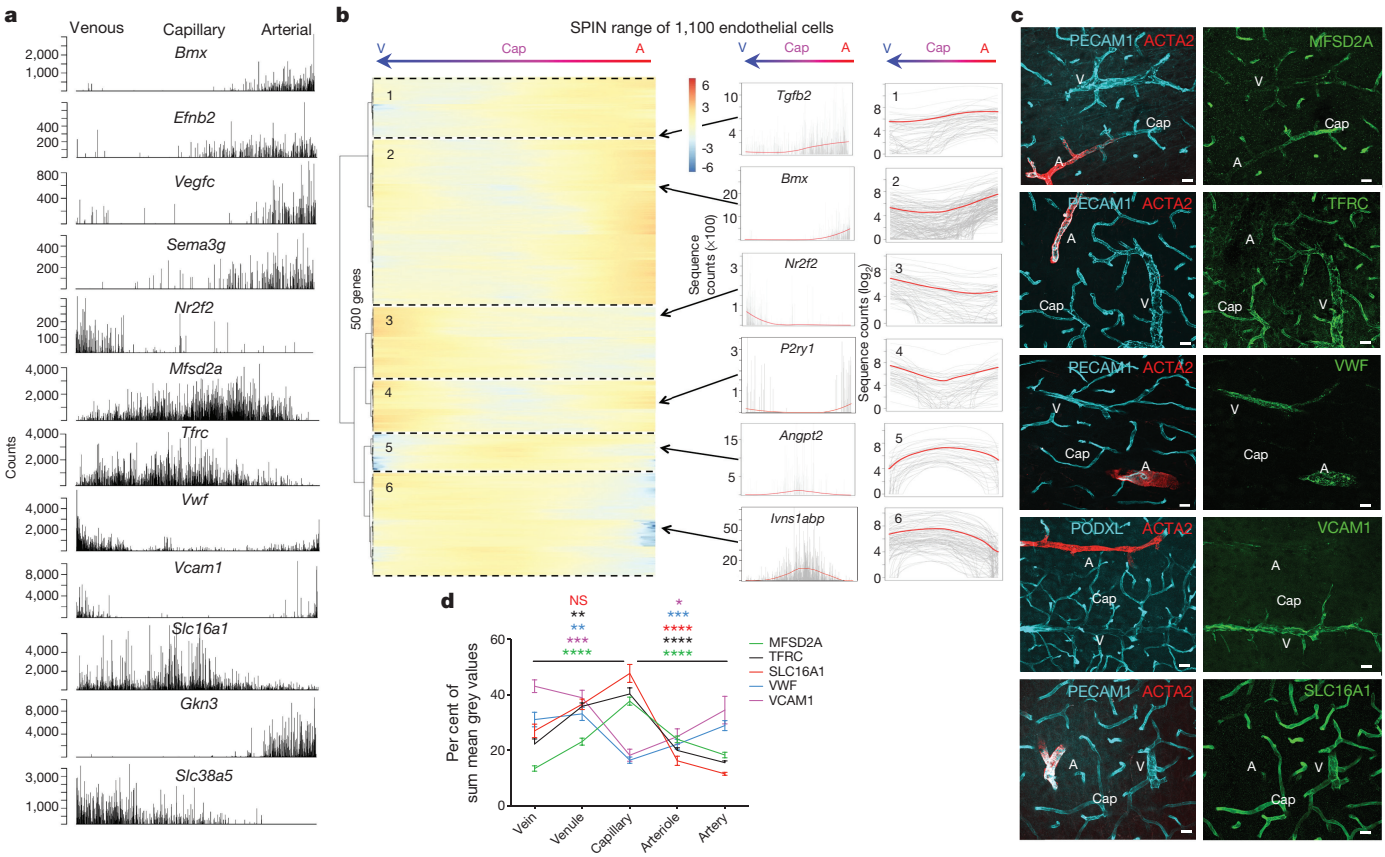


Figure 2 | Endothelial zonation. a, Zonal expression of transcripts across endothelial cells sorted by SPIN. **b**, Left, heat map of relative gene expression (red, high; blue, low) reveals six major patterns: artery (A; 1 and 2), vein (V; 3), artery and vein (A+V; 4), capillary (Cap; 5) and vein and capillary (V+Cap; 6). Middle, examples of individual genes. Arrows mark position in heat map. Grey bars show individual cells and red curves average expression. Right, curve plots for each group, average in red. **c**, Immunofluorescent staining of cerebral cortex for indicated markers. **d**, Quantification of immunofluorescent results. For each protein the number of mice and of image fields was MFSD2A 4, 20; TFRC 5, 36; SLC16A1 4, 21; VWF 4, 20; VCAM1 3, 20, respectively. All proteins show statistically significant zonation (Kruskal–Wallis, $P < 0.0001$). Dunn's multiple comparison test demonstrates statistically significant differences between vein and capillary and between artery and capillary (****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$, NS, not significant). Scale bars, 20 μm.

the transition from aaSMC to pericyte takes place between one cell and the next at this location (Extended Data Fig. 6f). These data indicate that mural cell arteriovenous zonation is punctuated into two separated phenotypic continua, in contrast to the single arteriovenous continuum of endothelial cells.

The zonal distribution of transcription factors and transporters in mural cells (Extended Data Fig. 7c, d) implies that there are gene regulatory and physiological differences between mural cell categories, and the numerous pericyte-enriched novel transcription factors will be interesting targets for future analysis. Notably, aSMCs were enriched for expression of immediate early genes (IEGs), including *Fos*, *Fosb*, *Jun*, *Junb* and *Egr1* (Extended Data Fig. 7e), which may reflect a rapid (within minutes) response of aSMCs to removal of natural physical conditions such as high shear and pressure. It was recently reported that the dissociation of muscle stem cells induces an IEG response[23], demonstrating that this may be a common confounder in single cell analysis.

Brain pericytes molecularly defined

The 1,088 transcriptomes of the PC cluster provide an opportunity to molecularly define the brain pericyte. Closer analysis revealed that the vast majority (about 99%) of pericytes were free from contamination by any of the other analysed cell classes. The remaining roughly 1% of the pericytes contained canonical endothelial transcripts at 5–10% of their average levels in endothelial cells (Extended Data Fig. 9a). We assume that these pericytes were contaminated by endothelial cell fragments remaining attached to the pericytes during isolation. Using protocols in which cell dissociation conditions were deliberately relaxed, we found abundant endothelial–pericyte transcript contamination (data not shown). This type of contamination is also noticeable in previously published data[17,24].

The non-endothelial-contaminated population of pericytes did not form obvious subclusters, and a manual gene search of the BackSPIN data did not reveal the existence of brain pericyte subtypes. A distinguishing criterion for the PC cluster was absent (or extremely low) *Acta2* expression. The lack of ACTA2 signal in venules suggests that capillaries and venules harbour transcriptionally indistinguishable pericytes. The distribution of *abcc9* in zebrafish (Fig. 3b and Extended Data Fig. 8b) further supports this notion.

The lack of heterogeneity among pericytes in the brain prompted us to investigate possible heterogeneity between organs. For a first insight into pericyte organotypicity, we identified pericytes in the adult lung of *Pdgfrb*-GFP mice as strongly GFP-positive cells located on alveolar capillaries (Fig. 4a) and isolated these cells for scRNA-seq. Altogether, about 1,500 lung cells were sequenced and clustered by BackSPIN. In addition to pericytes, we isolated and annotated endothelial cells, SMCs and fibroblast- or cartilage-like cell types (which are not further discussed here) following the same strategy as for the brain cells. We identified a distinct cluster of lung pericytes based on the presence of the canonical pericyte markers *Pdgfrb*, *Cspg4*, and *Des* and the simultaneous absence of SMC markers (*Acta2* and *Tagln*) and fibroblast markers (*Pdgfra*, *Lum* and *Dcn*) (Extended Data Fig. 9b). In addition,

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

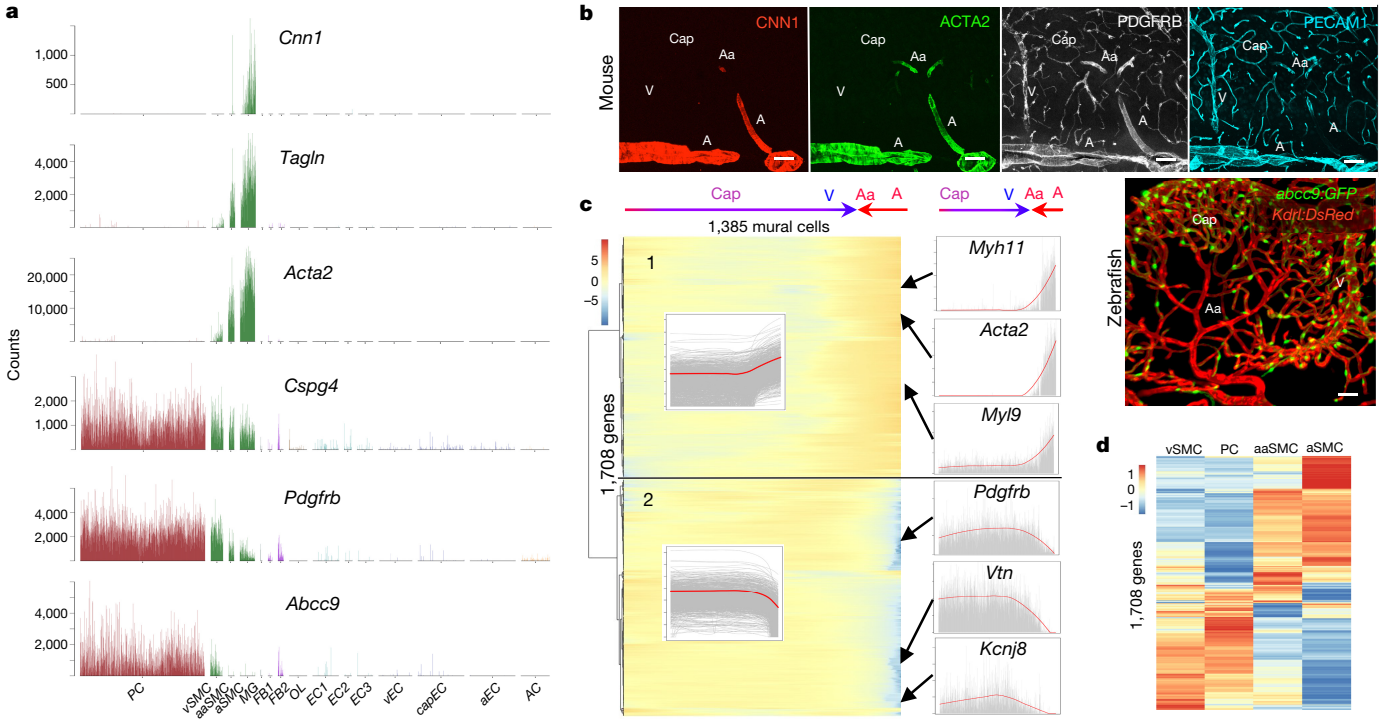


Figure 3 | Mural cell zonation. a, Bar plots of selected markers (see Extended Data Fig. 1 for cell type abbreviations). **b**, Immunofluorescence (mice, $n = 3$) and reporter gene (zebrafish) expression of selected genes and proteins from **a** (see also Extended Data Fig. 7a, b). **c**, Cells from the PC, vSMC, aaSMC and aSMC clusters re-sorted into a SPIN range (*x*-axis) and hierarchical clustering (*y*-axis) shows two major patterns with preferences for aaSMC and aSMC (1) and PC and vSMC (2). Right: three examples from each group. Aa, arterioles. **d**, Expression heat map (red, high; blue, low) with cell clusters in anatomical order. All images from cerebral cortex. Scale bars, 50 μm.

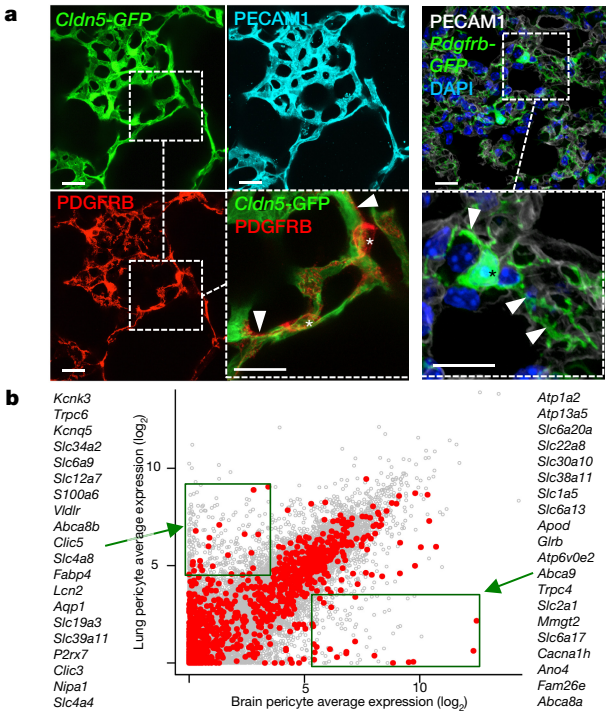


Figure 4 | Pericyte organotypicity. a, *Cldn5*-GFP and PECAM1 immunofluorescence mark endothelial cells, and *Pdgfrb*-GFP and PDGFRB immunofluorescence mark pericytes in mouse lung. Insets show pericytes at high magnification. Asterisks, cell bodies; arrowheads, pericyte processes. Scale bars, 20 μm. **b**, Transporters (red) in brain and lung pericytes. The 20 most differentially expressed transporters are indicated for each cell type.

lung pericytes shared the following markers with brain pericytes: *Vtn*, *Higd1b*, *S1pr3*, *Mcam*, *Ifitm1*, *Baiap3* and *Ehd3* (Extended Data Fig. 9b and http://betsholtzlab.org/VascularSingleCells/database.html). By extending the comparison to all genes, we identified substantial organotypic differences between brain and lung pericytes. For example, the commonly used marker for brain pericytes CD13 (encoded by *Anpep*) was not expressed by lung pericytes (Extended Data Fig. 9b). Differences were also apparent through *t*-distributed stochastic neighbour embedding (*t*-SNE) analysis, which clearly separated brain and lung pericytes (Extended Data Fig. 9c). A noticeable difference between brain and lung pericytes involved the SLC, ABC and ATP transporters, many of which (for example, *Atp13a5*) were abundant in brain pericytes (Extended Data Fig. 9d, e) but low or absent in lung pericytes (Fig. 4b). The gene ontology (GO) term 'transmembrane transporter activity' was associated with genes that were overexpressed in brain pericytes when compared with lung pericytes (data not shown). Beyond providing evidence for organotypic specialization of pericytes, this observation suggests that brain pericytes are directly involved in molecular transport at the BBB.

Perivascular fibroblast-like cells

We found a population of *Pdgfra*-driven histone 2B-fused GFP (H2BGFP)-positive cells (hereinafter referred to as brain fibroblast-like cells) in close association with arteries, arterioles, veins and venules. These cells were non-overlapping with endothelial and mural cells and were located between the vessel wall and the AQP4- and GFAP-positive astrocyte end-feet (Fig. 5a, b and Extended Data Fig. 10a, b). By contrast, *Pdgfra*-H2BGFP-positive oligodendrocyte progenitors, which are abundant in the brain parenchyma and occasionally apposed to blood vessels, were always located outside the AQP4-positive astrocyte end-feet (Fig. 5b, Extended Data Fig. 10b and Supplementary Video 1). Thus, the fibroblast-like cells reside within the perivascular Virchow–Robin space, which has been implicated in fluid exchange between the

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

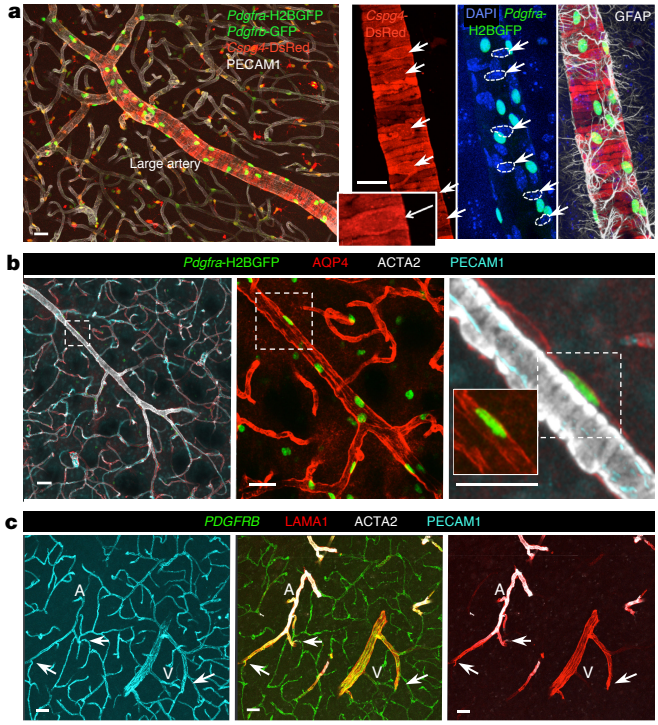


Figure 5 | Brain fibroblast-like cells. a, *Pdgfra*-H2BGFP-positive cells along arteries in cerebral cortex. Arrows indicate SMC nuclei. **b**, *Pdgfra*-H2BGFP-positive cells in arterioles are located between SMCs and astrocyte end-feet. **c**, LAMA1-positive ECM sleeve surrounds arteries, arterioles, veins and venules, but not capillaries. White arrows indicate transition from LAMA1-positive to LAMA1-negative vessel. Observations from 5 (**a**, **c**) and 2 (**b**) mice. Scale bars, 20 μm.

brain and the cerebrospinal fluid[25]. In contrast to mural cells, which are firmly embedded in the basement membrane, the fibroblast-like cells were only loosely adhered to the vessels and required relaxed digestion conditions to remain attached. Both BackSPIN and *t*-SNE analysis provided distinct clustering of brain fibroblast-like cells from all other vascular cell types (Fig. 1b and Extended Data Figs 1j, k, 9c). Subtypes of these cells could be distinguished (FB1 and FB2), but are not discussed further here.

We assigned the 50 most specific markers for the FB, PC, a/aaSMC and EC clusters by correlation with *Pdgfra*, *Abcc9*, *Acta2* and *Cldn5*, respectively (Extended Data Fig. 11a). Many of the FB-specific transcripts encode structural components, modifiers or receptors for the extracellular matrix (ECM) (Extended Data Fig. 11b), including fibrillar (for example, *Col1a1*, *Col1a2*, *Col3a1*, *Col5a1*, *Col5a2* and *Col5a3*) and non-fibrillar collagens (for example, *Col6a1*, *Col6a2*, *Col6a3*, *Col8a1*, *Col8a2*, *Col11a1*, *Col12a1*, *Col13a1*, *Col15a1*, *Col16a1*, *Col23a1* and *Col26a1*) (Extended Data Fig. 11c, d), collagen-modifying enzymes (for example, lysyl oxidases) and proteins involved in collagen fibril spacing, such as the small leucin-rich proteoglycans lumican (*Lum*) and decorin (*Dcn*) (Extended Data Fig. 11e, f). GO terms associated with ECM, collagen, cell adhesion and migration were significantly ($P < 1 \times 10^{-10}$) associated with the fibroblast-like cells; the other vascular cell types (PC, a/aaSMC and EC) instead showed enrichment for terms associated with cation transporter, muscle contraction and cell junction, respectively (Extended Data Fig. 11a). Of the 50 brain FB-specific transcripts, 45 were also expressed by lung fibroblasts (Extended Data Fig. 11b) pointing to a strong resemblance between these cells. However, the brain FB cells also expressed the endothelial marker *Cdh5*, the epithelial marker *Lama1* and the mesothelial marker *Efemp1*, suggesting that they may have mixed mesenchymal and epithelial characteristics (Extended Data Fig. 11b and http://betsholtzlab.org/VascularSingleCells/database.html).

The specific expression of *Lama1* allowed us to further study the anatomical distribution of brain fibroblasts. We found that LAMA1 antibodies decorate a continuous ECM sleeve located around arteries, arterioles, venules and veins, but not around capillaries, matching the perivascular distribution of *Pdgfra*-H2BGFP-positive cells (Fig. 5c and Extended Data Fig. 10c).

Discussion

Here, we have provided molecular definitions of the principal classes of brain vascular cells through their genome-wide quantitative transcriptomes. We have also unravelled a molecular blueprint for zonation along the arteriovenous axis, and illustrated how endothelial and mural cells differ in this regard. Single-cell analysis was crucial for this information, which would have been lost in the average profiles provided through bulk RNA sequencing. The examples of transcription factors and transporters illustrate that the zonal distribution of any class of proteins can now be assessed at http://betsholtzlab.org/VascularSingleCells/database.html. Our data also expand by around 100-fold the number of brain arteriovenous markers and pinpoint several arteriovenous landmarks that can be applied for vessel type identification in the brain. The cellular correlates of these landmarks are schematically summarized in Extended Data Fig. 6a.

Despite their functional importance and implication in a number of processes, pericytes remain poorly defined. Their proposed roles in vessel stabilization[26], BBB formation[27,28], immune cell guidance[29], blood flow regulation[30], tissue repair[31] and pathological scarring[32] are all clouded by ambiguities regarding cell type identification[33–35]. The molecular definition of brain and lung pericytes reveals several new markers common for these two types of pericyte, but also an extensive organotypicity[36]. It should be remembered that the cell type definitions presented herein are applicable only to the normal adult mouse brain and lung, and that heterogeneity may be expected in development and disease, topics that can now be systematically addressed.

The molecular identification of fibroblast-like cells and elucidation of their anatomical location provide insights into the cellular anatomy of the brain vasculature. The expression of markers on these cells matches that of scar-forming cells in spinal cord injury[32,33] and recently described perivascular cells with proposed pericyte-like characteristics[18]. Our analysis instead reveals a resemblance to lung fibroblasts combined with the expression of certain epithelial, endothelial and mesothelial markers. The location of these cells within a perivascular space that has been implicated in the glymphatic system[37] warrants investigations into their involvement in brain fluid homeostasis.

While it will now be important to map the transcriptional changes that occur in vascular cells during disease, the data available herein and at http://betsholtzlab.org/VascularSingleCells/database.html may already provide clues to cell types involved in brain diseases that feature neurovascular dysfunction, such as Alzheimer's disease, certain monogenic neurological diseases and brain tumours. For these conditions, our data pinpoint pericytes and fibroblast-like cells as common cellular culprits besides endothelial cells (Extended Data Fig. 12), suggesting that cellular involvement in neurovascular pathologies is more complex than anticipated.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 June 2017; accepted 10 January 2018.
Published online 14 February 2018.

1. Iadecola, C. The pathobiology of vascular dementia. *Neuron* **80,** 844–866 (2013).
2. Zhao, Z., Nelson, A. R., Betsholtz, C. & Zlokovic, B. V. Establishment and dysfunction of the blood-brain barrier. *Cell* **163,** 1064–1078 (2015).
3. Abbott, N. J., Rönnbäck, L. & Hansson, E. Astrocyte–endothelial interactions at the blood-brain barrier. *Nat. Rev. Neurosci.* **7,** 41–53 (2006).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

4. Zlokovic, B. V. The blood-brain barrier in health and chronic neurodegenerative disorders. *Neuron* **57,** 178–201 (2008).
5. Phoenix, T. N. *et al.* Medulloblastoma genotype dictates blood brain barrier phenotype. *Cancer Cell* **29,** 508–522 (2016).
6. de Lange, E. C. & Hammarlund-Udenaes, M. Translational aspects of blood-brain barrier transport and central nervous system effects of drugs: from discovery to patients. *Clin. Pharmacol. Ther.* **97,** 380–394 (2015).
7. Majno, G. & Palade, G. E. Studies on inflammation. 1. The effect of histamine and serotonin on vascular permeability: an electron microscopic study. *J. Biophys. Biochem. Cytol.* **11,** 571–605 (1961).
8. Simionescu, M., Simionescu, N. & Palade, G. E. Segmental differentiations of cell junctions in the vascular endothelium. The microvasculature. *J. Cell Biol.* **67,** 863–885 (1975).
9. Ekman, N. *et al. Bmx* tyrosine kinase is specifically expressed in the endocardium and the endothelium of large arteries. *Circulation* **96,** 1729–1732 (1997).
10. Wang, H. U., Chen, Z. F. & Anderson, D. J. Molecular distinction and angiogenic interaction between embryonic arteries and veins revealed by ephrin-B2 and its receptor Eph-B4. *Cell* **93,** 741–753 (1998).
11. Hogan, B. M. *et al.* Vegfc/Flt4 signalling is suppressed by Dll4 in developing zebrafish intersegmental arteries. *Development* **136,** 4001–4009 (2009).
12. Kutschera, S. *et al.* Differential endothelial transcriptomics identifies semaphorin 3G as a vascular class 3 semaphorin. *Arterioscler. Thromb. Vasc. Biol.* **31,** 151–159 (2011).
13. Hirashima, M. & Suda, T. Differentiation of arterial and venous endothelial cells and vascular morphogenesis. *Endothelium* **13,** 137–145 (2006).
14. Armulik, A., Genové, G. & Betsholtz, C. Pericytes: developmental, physiological, and pathological perspectives, problems, and promises. *Dev. Cell* **21,** 193–215 (2011).
15. Attwell, D., Mishra, A., Hall, C. N., O'Farrell, F. M. & Dalkara, T. What is a pericyte? *J. Cereb. Blood Flow Metab.* **36,** 451–455 (2016).
16. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9,** 171–181 (2014).
17. Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347,** 1138–1142 (2015).
18. Marques, S. *et al.* Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352,** 1326–1329 (2016).
19. Goldmann, T. *et al.* Origin, fate and dynamics of macrophages at central nervous system interfaces. *Nat. Immunol.* **17,** 797–805 (2016).
20. Halpern, K. B. *et al.* Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542,** 352–356 (2017).
21. Nguyen, L. N. *et al.* Mfsd2a is a transporter for the essential omega-3 fatty acid docosahexaenoic acid. *Nature* **509,** 503–506 (2014).
22. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445,** 168–176 (2007).
23. van den Brink, S. C. *et al.* Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14,** 935–936 (2017).
24. Zhang, Y. *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34,** 11929–11947 (2014).
25. Iliff, J. J. *et al.* A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β. *Sci. Transl. Med.* **4,** 147ra111 (2012).
26. Lindahl, P., Johansson, B. R., Levéen, P. & Betsholtz, C. Pericyte loss and microaneurysm formation in PDGF-B-deficient mice. *Science* **277,** 242–245 (1997).
27. Armulik, A. *et al.* Pericytes regulate the blood-brain barrier. *Nature* **468,** 557–561 (2010).
28. Daneman, R., Zhou, L., Kebede, A. A. & Barres, B. A. Pericytes are required for blood-brain barrier integrity during embryogenesis. *Nature* **468,** 562–566 (2010).
29. Proebstl, D. *et al.* Pericytes support neutrophil subendothelial cell crawling and breaching of venular walls *in vivo*. *J. Exp. Med*. **209,** 1219–1234 (2012).
30. Hall, C. N. *et al.* Capillary pericytes regulate cerebral blood flow in health and disease. *Nature* **508,** 55–60 (2014).
31. Crisan, M. *et al.* A perivascular origin for mesenchymal stem cells in multiple human organs. *Cell Stem Cell* **3,** 301–313 (2008).
32. Göritz, C. *et al.* A pericyte origin of spinal cord scar tissue. *Science* **333,** 238–242 (2011).
33. Soderblom, C. *et al.* Perivascular fibroblasts form the fibrotic scar after contusive spinal cord injury. *J. Neurosci.* **33,** 13882–13887 (2013).
34. Hill, R. A. *et al.* Regional blood flow in the normal and ischemic brain is controlled by arteriolar smooth muscle cell contractility and not by capillary pericytes. *Neuron* **87,** 95–110 (2015).
35. Guimaraes-Camboa, N. *et al.* Pericytes of multiple organs do not behave as mesenchymal stem cells *in vivo*. *Cell Stem Cell* **20,** 345–359 (2017).
36. Nolan, D. J. *et al.* Molecular signatures of tissue-specific microvascular endothelial cell heterogeneity in organ maintenance and regeneration. *Dev. Cell* **26,** 204–219 (2013).
37. Xie, L. *et al.* Sleep drives metabolite clearance from the adult brain. *Science* **342,** 373–377 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was supported by AstraZeneca AB (C.B., U.L.), the Swedish Research Council (C.B.: 2015-00550; U.L.: K2014-64X-20097-09-5), the European Research Council (C.B.: AdG294556), the Leducq Foundation (C.B., A.K.: 14CVD02), Swedish Cancer Society (C.B.:150735; U.L.:CAN 2016/271), Knut and Alice Wallenberg Foundation (C.B.: 2015.0030), Hjärnfonden (U.L.), Swiss National Science Foundation (A.K.: 31003A_159514/1) and the Synapsis Foundation (A.K.). We thank C. Olsson, H. Leksell, P. Peterson, J. Chmielniakova, K. Gaengel, BioVis (Uppsala), Center for Microscopy and Image Analysis (Zurich) and Eukaryotic Single Cell Genomics facility (Science for Life Laboratory) for technical help, and K. Alitalo, P. Soriano, D. Silver and L. Sorokin for reagents.

Author Contributions M.V., L.H., N.M., U.L. and C.B. conceived and designed the project; M.V., M.A.M., J.A., K.A., F.D.G., K.N., T.L., B.L., E.R., L.G., Y.Z., M.R., A.K. and C.B. performed experiments; L.H. performed bioinformatic analysis; L.H. and Y.S. constructed the online database; C.B., L.H. and M.V. analysed the bioinformatic data; C.B. and U.L. wrote the manuscript with substantial input from M.V., L.H., M.A.M., J.A. and K.A. All authors reviewed the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to C.B. (christer.betsholtz@ki.se).

Reviewer Information *Nature* thanks D. McDonald and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Mice. The following mouse strains were used: *C57Bl6* (The Jackson Laboratory, *C57BL6/J*), *Bmx*^{Lacz} (*Bmx*^{tm1Ali})[38], *Cspg4*-DsRed (The Jackson Laboratory, Tg(*Cspg4*-DsRed.T1)1Akik/J, *Pdgfrb*-GFP (Gensat.org. Tg(*Pdgfrb*-eGFP) JN169Gsat/Mmucd)[39], *Pdgfra*-H2BGFP (*Pdgfra*$^{tm11(EGFP)Sor}$, gift from P. Soriano), *Cldn5*-GFP (Tg(*Cldn5*-GFP)Cbet/U), *SM22*-Cre (Tg(*Tagln*-cre)1Her/J and *R26-stop*-tdTomato (*B6;129S6-Gt(ROSA)26Sor*$^{tm14(CAG-tdTomato)Hze}$). All mice were back-crossed on a *C57BL6/J* genetic background. Adult mice of either sex aged 10–19 weeks were used for all experiments. To generate the brain single-cell dataset we used the following mice: two mice of *Pdgfrb*-GFP;*Cspg4*-DsRed genotype, two mice of *Cldn5*-GFP;*Cspg4*-DsRed genotype, three mice of *Pdgfra*-H2BGFP genotype and three mice of *SM22-Cre;R26-stop-tdTomato* genotype. The lung single-cell dataset was generated from two *Pdgfrb*-GFP;*Cspg4*-DsRed mice, one *Pdgfrb*-GFP mouse and two *Cldn5*-GFP;*Cspg4*-DsRed mice. No statistical method was used to predetermine sample size. Animal experiment protocols were approved by the Uppsala Ethical Committee on Animal Research (permit numbers C224/12 and C115/15) and the Stockholm North Animal Ethics committee (Stockholms Norra Djurförsöksetiska Nämnd), permit N150/14. All animal experiments were carried out in accordance with their guidelines.

Single cell isolation. The mice were killed by cervical dislocation, after which the brain was surgically removed and placed in ice-cold Dulbecco's modified Eagle's medium (DMEM, ThermoFisher Scientific) supplemented with penicillin/streptomycin (ThermoFisher Scientific cat# 15140122). For all brain single-cell isolations, the entire brain, without olfactory bulbs, was used.

Brain cell isolation. Brain mural cells (*Pdgfrb*-GFP/*Cspg4*-DsRed positive) and brain endothelial cells (*Cldn5*-GFP positive) were isolated with a modified version of the Neural Tissue Dissociation kit (P) (Myltenyi Biotec, Cat# 130-092-628). The brain tissue was mechanically and enzymatically dissociated before removal of myelin, which is detrimental for efficient sorting. Myelin was removed using magnetic bead separation (Myelin Removal Beads II, Myltenyi Biotec, cat# 130-096-733) according to the manufacturer's protocol. The final cell pellet was resuspended in 500–1,000 µl FACS buffer (DMEM without phenol red (ThermoFisher Scientific), supplemented with 2% fetal bovine serum (FBS, ThermoFisher Scientific). For a detailed protocol, see Protocol Exchange[40].

Perivascular single-cell isolation. In order to enrich for perivascular *Pdgfra*-H2BGFP-positive cells from the brain before sorting, the brain vasculature was isolated with Dynabeads (Dynabeads Sheep Anti-Rat IgG, ThermoFisher Scientific, cat# 11035) coupled to rat anti-mouse CD31 (BD Pharmingen, cat# 553370) before the tissue was digested into single cells. A detailed description of the protocol can be found in Protocol Exchange[41]. After microvascular isolation (and before single-cell dissociation), the microvascular fragments were microscopically inspected to verify the presence of microvascular fragments with *Pdgfra*-H2BGFP-positive cells still attached (Leica Inverted microscope, Leica Microsystems). The final pellet was resuspended in 500 µl FACS buffer.

Lung single-cell isolation. To isolate single lung cells, mice were killed as described above and the lungs were removed. Single cells were isolated using a combination of mechanical and enzymatic dissociation. The primary enzymes used were collagenase type 2 (ThermoFisher Scientific, cat#C6885) and enzymes 3 and 4 from the Neural Tissue Dissociation kit (P), described above. For detailed instructions, see Protocol Exchange[42]. The obtained cell pellet was resuspended in 500 µl FACS buffer.

Fluorescence-activated cell sorting (FACS). Prior to FACS, all single-cell suspensions were strained over a 35-µm mesh (Cell-Strainer capped tubes, Corning, cat# 352235). Note that FACS was used merely to enrich and capture the cells of interest, and not for identification. For this reason, all FACS settings were set very broadly to avoid FACS-induced bias. All cell suspensions were sorted on a BD FACSAria III (BD Biosciences) at the BioVis core facility of the Department of Immunology, Genetics and Pathology (IGP), Uppsala University. A 100-µm nozzle and 20-psi PBS sheet fluid pressure were used. First, cells were selected using forward scatter area/side scatter area (FSC-A/SSC-A, linear scale) with a very wide gate setting, to accommodate all possible cell sizes and morphologies with as little bias as possible. For the same reason, no doublet discrimination was implemented, since this might favour the enrichment of cells with a uniform, round morphology. Next, fluorescent events based on the parent FSC-A/SSC-A gate were selected: GFP was excited with a 488-nm laser, and emission was detected through a 530/30 filter, while DsRed or TdTomato was exited with a 561-nm laser and emission detected through a 582/15 filter. As controls, wild-type *C57Bl6/J* mice lacking fluorescent reporters were used. Single cells, positive for GFP alone (*Cldn5*-GFP;*Cspg4*-DsRed and *Pdgfra*-H2BGFP mice), positive for both GFP and DsRed (*Pdgfrb*-GFP;*Cspg4*-DsRed mice) or positive for TdTomato (*SM22-Cre;R26-stop-tdTomato*) were sorted straight into 2.3 µl lysis buffer pre-dispensed into a 384-well plate. For all sorts, the single cell mask was used. Prior to cell collection, special care was taken to ensure that cells were sorted directly into the lysis buffer without touching the well wall: 1) The waste collection bin was moved to the right as much as possible (while still catching the waste stream) so that the sort angle could be reduced to 8–10 units. 2) Sort accuracy was optimized by sorting 100 BD FACS Accudrop beads (BD Biosciences) per well on the lid of an empty 384-well plate, followed by visual inspection of sort drop location. If needed, the plate position was recalibrated to ensure the best sort accuracy. All FACS data were analysed using FlowJo v10.1 (FlowJo, LLC) and FACSDiva v8.0.2 (BD Biosciences).

Smart-seq2 library prep and sequencing. Single-cell libraries were prepared as described previously[43], with the following specifications: 0.0025 µl of a 1:40,000 diluted ERCC spike-in concentration stock was used, and all cDNA was amplified with 22 PCR cycles before QC control with a Bioanalyzer (Agilent Biosystems). The libraries were sequenced on a HiSeq2500 at the National Genomics Infrastructure (NGI), Science for Life Laboratory, Sweden, with single 50-bp reads (dual indexing reads). All single-cell transcriptome data were generated at the Eukaryotic Single-cell Genomics facility at Science for Life Laboratory in Stockholm, Sweden.

Antibodies and immunofluorescence staining. Mice under full anaesthesia were euthanized by transcardial perfusion with Hanks balanced salt solution (HBSS, GIBCO, cat# 14025092) followed by 4% buffered formaldehyde (Histolab, cat# 02178). Brains were removed and postfixed in 4% buffered formaldehyde for 4 h (or overnight) at 4 °C. Sagittal and coronal vibratome sections (50–150 µm) were incubated in blocking/permeabilization solution (1% bovine serum albumin, 2.5% donkey serum, 0.5% Triton X-100 in PBS) overnight at 4 °C, followed by incubation in primary antibody solution for two nights at 4 °C, and subsequently in secondary antibody solution, overnight at 4 °C. Sections were mounted in ProLong Gold Antifade mountant (Life Technologies, cat# P36930). Images were taken with a Leica TCS SP8 confocal microscope (Leica Microsystems). An overview of the antibodies used is presented in Supplementary Information Table 4. All images were acquired with a Leica TCS SP8 confocal microscope (Leica Microsystems) and Leica Application Suite 3.2.1.9702. Images were adjusted for brightness and contrast using Leica Application Suite 3.2.1.9702, Fiji v1.51 s and Adobe PhotoShop CC 2015. All confocal images are represented as maximum intensity projections unless stated otherwise (Leica Application Suite 3.2.1.9702 and Fiji v1.51 s).

X-gal staining. *Bmx*^{LacZ} knock-in reporter mice[38] were used to visualize *Bmx* expression in mouse brains with X-gal staining. First, the mice were perfused with PBS, followed by 4% PFA. Dissected brains were postfixed in 4% PFA for 60 min at 4 °C, washed and cut with a vibratome into 50-µm-thick coronal sections. The brain sections were washed 3 times for 20 min with 0.1 M phosphate buffer containing 0.02% NP-40, 0.01% deoxycholate and 2 mM $MgCl_2$. Next, the sections were incubated in staining solution containing 1 mg/ml X-gal, 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2 mM $MgCl_2$, 0.02% NP-40, 0.01% deoxycholate and 0.1 M phosphate buffer for 21 h at 37 °C. The brain slices were washed extensively with PBS, blocked for 7 h at 4 °C in 1% BSA, 0.5% Triton X-100 in PBS, before adding primary antibody solution. The rest of the protocol was continued as described in the previous paragraph.

RNA *in situ* hybridization. RNA *in situ* hybridization was performed using the RNAscope technology (Advanced Cell Diagnostics) following the manufacturer's protocol with minor modifications. In brief, fresh-frozen brains were cut into 20-µm sagittal sections and mounted on SuperFrost Plus glass slides. After dehydration, slides were subjected to RNAscope Multiplex Fluorescent Assay. First, slides were incubated in Pretreat 3 for 30 min at room temperature. After that, RNAscope probes, *Gkn3-C1*, *Slc38a5-C1*, *Acta2-C2* and *Pecam1-C3*, were hybridized for 2 h at 40 °C and the remainder of the assay protocol was implemented. The fluorescent signal emanating from RNA probes was visualized and captured using a Leica TCS SP8 confocal microscope (Leica Microsystems). All *in situ* hybridization images presented are 2D maximum intensity projections of ~4-µm *z*-stacks obtained from cerebral cortex.

Quantifications. *Vessel diameter.* Vessel diameter and length were measured in 75–150-µm vibratome brain sections obtained from *C57Bl6/J* and *Pdgfra*-H2BGFP mice. Four to seven areas from cerebral cortex in three animals were analysed. Blood vessels were visualized by CD31 and perivascular cells by TAGLN, ACTA2, CNN1 or LAMA1 immunostaining. *Z*-stacks were captured using a Leica TCS SP8 confocal microscope (Leica Microsystems). Vessel diameter and length were measured manually using Leica Application Suite 3.2.1.9702 software.

Endothelial zonation. Quantifications were performed on 8-bit confocal images obtained from samples visualized by immunohistochemistry or RNA *in situ* hybridization (as described above). For quantifying immunohistochemical images, the ROI manager of the Fiji software was used to mark and measure the mean grey values of the vascular segments of interest (vein, venule, capillary, arteriole, and artery). The sum of all mean grey values per field was set as 100% and the percentage of mean grey values per vascular segment was calculated for each image. A Kruskal–Wallis test with Dunn's multiple comparison test was performed to demonstrate statistically significant differences in protein expression along the arteriovenous axis. The data were presented using GraphPad Prism 6 software. For

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

the quantification of RNA *in situ* hybridization images, the ROI manager of the Fiji software was used to mark and measure the mean grey values of the vascular segments of interest (vein, artery) and presented using GraphPad Prism 6 software. An unpaired *t*-test with Welsh correction was used to demonstrate statistical significance in gene expression between arteries and veins.

Zebrafish. *BAC recombination.* The pRedET plasmid (GeneBridge) was introduced into *Escherichia coli* containing a CH211-58C15 BAC clone encoding the *abcc9* gene (BacPAC Resources) by electroporation (1,800 V, 25 μF, 200 Ω). Next, two *Tol2* long terminal repeats in opposing directions flanking an ampicillin resistance cassette were amplified by PCR using Tol2_amp as a template and were inserted into the BAC vector backbone. Then, cDNA encoding IRES-Gal4FF together with a kanamycin resistance cassette (IRES_Gal4FF_KanR) were amplified by PCR using pCS2_mCherry_IRES_Gal4FF_KanR plasmid as a template and inserted at the C terminus of the *abcc9* gene (ENSDART00000079987.4). Primers to amplify the IRES_Gal4FF_KanR PCR product were as follows:

5′- atggagcaggaggacggcctgtttgcatcttttgtcaaagccgacatgtaGATGAGTAAC TTGTACAAAG-3′ and 5′- aaaatggcttttattgatctgttaaggccaaaagtggtgtaaagtgggga TCAGAAGAACTCGTCAAGAAGGCG-3′ (lowercase, homology arm to BAC vector; uppercase, primer binding site to the template plasmid).

*Plasmids***.** The Tol2_amp and pCS2_Gal4FF_KanR vector for BAC recombination were kindly provided by S. Schulte-Merker (University of Münster, Germany). To construct the plasmid encoding pCS2-mCherry_IRES_Gal4FF_KanR, the pCS2_Gal4FF_KnaR vector had the following cDNA inserted: a PCR-amplified IRES using p3E-IRES-nlsEGFP in Tol2Kit[44] as a template and pmCherry-C1 (Clontech, Takara Bio Inc.).

Transgenic and mutant zebrafish lines. The Tol2 transposase mRNA was transcribed *in vitro* with SP6 RNA polymerase from a NotI-linearized pCS-TP vector using the mMESSAGE mMACHINE kit (Ambion). To generate the *TgBAC(abcc9:Gal4FF)^{ncv34}* zebrafish line, the corresponding BAC DNA (described above) was co-injected with Tol2 transposase mRNA into one-cell stage embryos of AB wild type. *Tg(UAS:GFP)* and *Tg(UAS:RFP)* fish lines[45] were kindly provided by K. Kawakami (National Institute of Genetics, Japan). *Tg(fli1a:Myr-mCherry)^{ncv1}*,*TgBAC(pdgfrb:Gal4FF)^{ncv}* and *TgBAC(tagln:EGFP)^{ncv25}* fish lines were previously described[44–47]. *Tg(kdrl:DsRed2)^{pd27}* was obtained from zebrafish international resource centre[48]. Throughout the text, all transgenic lines used in this study are simply described without their line numbers. For example, *TgBAC(abcc9:Gal4FF)^{ncv34}* is abbreviated to *TgBAC(abcc9:Gal4FF).*

Raw sequencing data processing. The samples were analysed by first demultiplexing the fastq files with deindexer (https://github.com/ws6/deindexer) using the Nextera index adapters and the 384-well layout. Individual fastq files were then mapped to the mouse reference genome (mm10) using the STAR aligner version 2.4.2a[49]. Two-pass alignment was chosen to have improved performance of *de novo* splice junction reads, filtered for only uniquely mapping reads. The expression values were computed per gene as described[50], using uniquely aligned reads and correcting for the uniquely alignable positions using MULTo[51]. The count matrix shows the individual counts aligning to each gene per cell. Cells with total aligned reads less than 100,000 were removed, and also outlier cells with maximum Spearman's correlation with other cells less than 0.3 were filtered out. After quality filtering, 3,186 brain cells and 1,504 lung cells were selected for further analysis. In order to compare the gene expression counts between different cells, the total gene counts for each cell were normalized to 500,000. Genes that were identified in less than three cells in the brain or lung were filtered out in the respective tissue analysis.

Cell type classification. In order to classify the cells into different cell types, the BackSPIN algorithm was used[17]. BackSPIN was run with the following parameters: -d 6 -g 3 -c 5, splitting the cells into six levels. The 3,186 brain cells were divided into 38 groups. BackSPIN iteratively splits the cells and the genes (assigning a proportion of the genes to each new branch), thus clusters that receive many genes expressed at background/noise levels continue to split at subsequent levels based on noise. Other clusters may instead stop splitting prematurely, because genes that were differentially expressed within the cluster had been assigned to another branch. The pericyte/vSMC cluster stopped splitting on the second level, because the SMC genes distinguishing pericytes from vSMC in this cluster had been assigned to another node which contained the aa/aSMCs and therefore expressed higher levels of SMC genes. For this reason, the pericyte cluster was re-run using BackSPIN including all genes, which led to clear separation of the pericytes and the vSMC. For other clusters, no other differences were found above noise by manual reading of all barplots, or, when such differences were found (such as in the oligodendrocyte and microglia clusters), they contained too few cells and were left without further splitting, or merged into an ancestral node. In the end, the 3,186 brain cells were classified into 15 clusters. Among them, we identified a small astrocyte cluster of six cells. To achieve a better representation of brain vessel-associated astrocyte gene expression profiles for comparison with the other vessel-associated cell types, we added a set of 250 pure brain astrocyte cells that were obtained in a parallel study in which cells were sorted from *SM22-Cre;R26-stop-tdTomato* mice. Although this project was focused on smooth muscle cells, it fortuitously also generated a significant number of *Aqp4*-positive astrocytes, apparently caused by ectopic reporter activation in this cell type. The full dataset, which will be published separately, was obtained and analysed using the same methodologies as described herein, including mouse genetic background. The 1,504 lung cells were divided into 46 groups by BackSPIN. After manual inspection, they were classified into 17 clusters. In order to better visualize expression differences between clusters, we created bar plots using average read count from all cells (even those with no expression).

Differential gene expression analysis. After cell type classification, the edgeR method was applied to compare the gene expression differences between different clusters[52]. In the pairwise comparisons of the three brain endothelial cell clusters (vEC, capEC and aEC), 1,798 genes were identified as being significantly differentially expressed (selection criteria: >twofold difference, $P < 0.05$ and represented in ≥10 cells in these three clusters). One-way ANOVA analysis was also applied to the 1,798 genes to compare their differential expression among the three brain endothelial cell clusters. Multiple test correction was implemented using the false discovery rate (FDR) method[53] in R. The top 500 genes with the smallest ANOVA *P* values were selected for further analysis in Fig. 2b. In the pairwise comparisons of the four brain mural cell clusters (aSMC, aaSMC, vSMC and PC), 1,708 genes were identified (selection criteria: >twofold difference, $P < 0.05$ and represented in >20% cells in the corresponding clusters).

***t*-SNE and heat map visualization.** The 2D *t*-SNE visualization was performed using the *t*-distributed stochastic neighbour embedding method for R software with default parameters (R tsne package version 0.1-3). The normalized gene count matrix for all genes was used as input. To visualize the gene expression level in cells on *t*-SNE map, the plotexptsne function in RaceID package (version 1.0) was adapted[54]. The heat map visualization was performed using the pheatmap packages (version 1.0.8) in R. The genes were clustered using Pearson correlation distance and the ward.D2 cluster method was used to build the cluster dendrogram.

Zonation analysis. The cells included in the zonation analysis were first sorted using the SPIN method, and then a smooth curve was fitted to the gene expression counts for each gene by loess function using the default parameters in R software.

Transcription factor and transporter analysis. To identify the transcription factors in the mural cells and endothelial cells, we made use of a previously reported list of 1,507 transcription factors on the mouse genome[55]. To identify transporter related genes, we made use of gene ontology annotation (www.geneontology.org) and all genes annotated as transporter activity (GO:0005215) or its child terms were extracted.

Gene ontology (GO) analysis. The GO analysis was performed using GOstat packages (version 2.40.0) in R software, with the associated annotation packages GO.db (version 3.4.0) and org.Mm.eg.db (version 3.4.0).

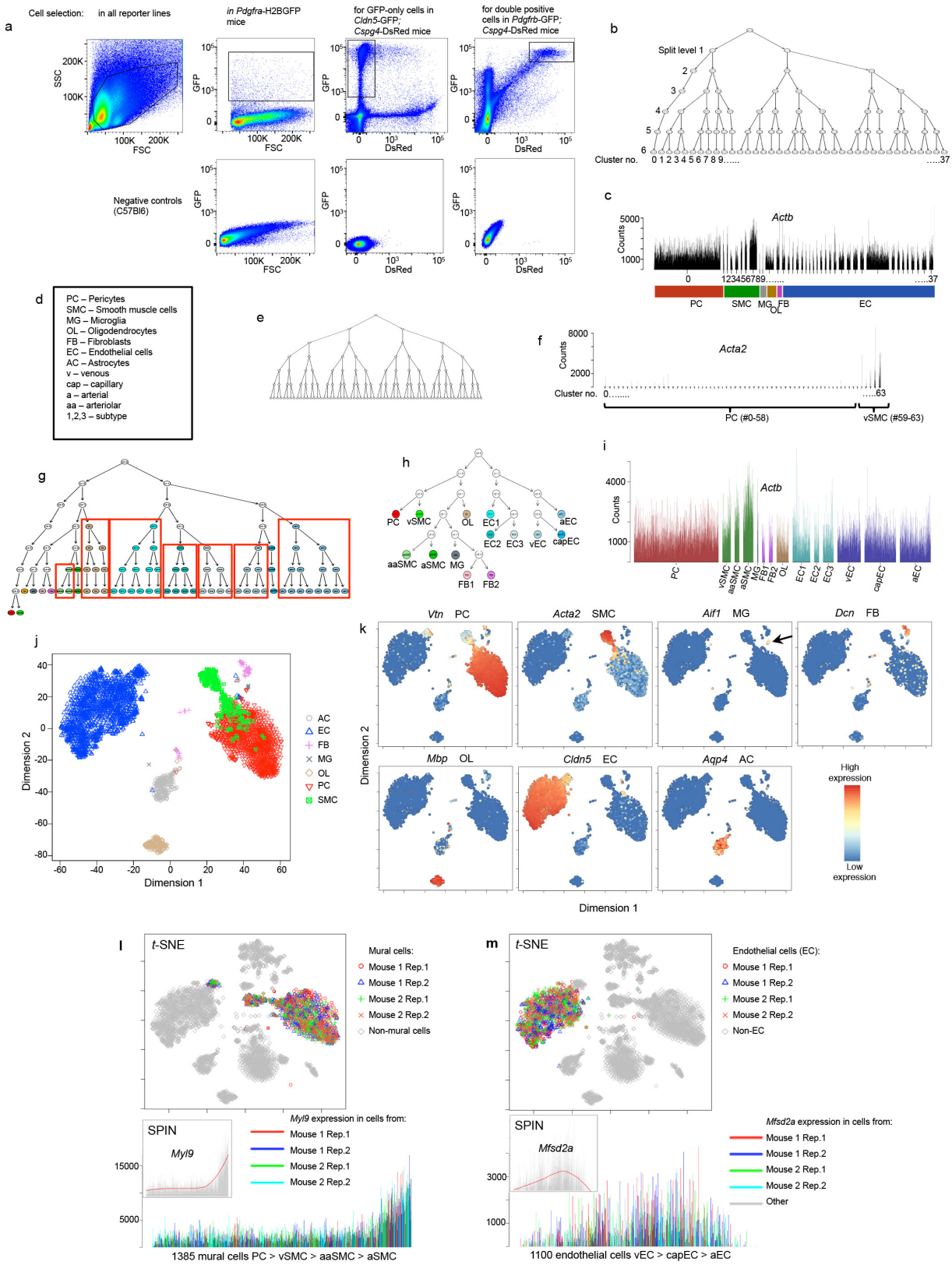
Data availability. All sequence data used in this study have been deposited in the NCBI Gene Expression Omnibus database and are accessible through accession numbers GSE98816, GSE99058 and GSE99235. The searchable database is freely available at http://betsholtzlab.org/VascularSingleCells/database.html.

38. Rajantie, I. *et al.* Bmx tyrosine kinase has a redundant function downstream of angiopoietin and vascular endothelial growth factor receptors in arterial endothelium. *Mol. Cell. Biol.* **21,** 4647–4655 (2001).
39. He, L. *et al.* Analysis of the brain mural cell transcriptome. *Sci. Rep.* **6,** 35108 (2016).
40. Vanlandewijck, M. *et al.* Primary isolation of vascular cells from murine brain for single cell sequencing. *Protoc. Exch.* https://doi.org/10.1038/protex.2017.159 (2018).
41. Vanlandewijck, M., Andrae, J., Gouveia, L. & Betsholtz, C. Isolation of vessel-associated pdgfra-H2BGFP positive cells from murine brain. *Protoc. Exch.* https://doi.org/10.1038/protex.2018.005 (2018).
42. Vanlandewijck, M., Andrae, J., Gouveia, L. & Betsholtz, C. Preparation of single cell suspensions from the adult mouse lung. *Protoc. Exch.* https://doi.org/10.1038/protex.2018.006 (2018).
43. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10,** 1096–1098 (2013).
44. Kwan, K. M. *et al.* The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev. Dyn.* **236,** 3088–3099 (2007).
45. Asakawa, K. *et al.* Genetic dissection of neural circuits by *Tol2* transposon-mediated Gal4 gene and enhancer trapping in zebrafish. *Proc. Natl Acad. Sci. USA* **105,** 1255–1260 (2008).
46. Fukuhara, S. *et al.* Visualizing the cell-cycle progression of endothelial cells in zebrafish. *Dev. Biol.* **393,** 10–23 (2014).
47. Ando, K. *et al.* Clarification of mural cell coverage of vascular endothelial cells by live imaging of zebrafish. *Development* **143,** 1328–1339 (2016).
48. Kikuchi, K. *et al.* Retinoic acid production by endocardium and epicardium is an injury response essential for zebrafish heart regeneration. *Dev. Cell* **20,** 397–404 (2011).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

49. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
50. Ramsköld, D., Kavak, E. & Sandberg, R. How to analyze gene expression using RNA-sequencing data. *Methods Mol. Biol*. **802,** 259–274 (2012).
51. Storvall, H., Ramskold, D. & Sandberg, R. Efficient and comprehensive representation of uniqueness for next-generation sequencing by minimum unique length analyses. *PLoS ONE* **8,** e53822 (2013).
52. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
53. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).
54. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525,** 251–255 (2015).
55. Forrest, A. R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507,** 462–470 (2014).
56. Mackic, J. B. *et al.* Human blood-brain barrier receptors for Alzheimer's amyloid-beta 1- 40. Asymmetrical binding, endocytosis, and transcytosis at the apical side of brain microvascular endothelial cell monolayer. *J. Clin. Invest.* **102,** 734–743 (1998).
57. Do, T. M. *et al.* Oatp1a4 and an L-thyroxine-sensitive transporter mediate the mouse blood-brain barrier transport of amyloid-β peptide. *J. Alzheimers Dis.* **36,** 555–561 (2013).
58. Deane, R. *et al.* LRP/amyloid beta-peptide interaction mediates differential brain efflux of Abeta isoforms. *Neuron* **43,** 333–344 (2004).
59. Bell, R. D. *et al.* SRF and myocardin regulate LRP-mediated amyloid-beta clearance in brain vascular cells. *Nat. Cell Biol.* **11,** 143–153 (2009).
60. Xiong, H. *et al.* ABCG2 is upregulated in Alzheimer's brain with cerebral amyloid angiopathy and may act as a gatekeeper at the blood-brain barrier for $A\beta_{1-40}$ peptides. *J. Neurosci.* **29,** 5463–5475 (2009).
61. Agarwal, S. *et al.* Active efflux of Dasatinib from the brain limits efficacy against murine glioblastoma: broad implications for the clinical use of molecularly targeted agents. *Mol. Cancer Ther.* **11,** 2183–2192 (2012).
62. Lam, F. C. *et al.* β-Amyloid efflux mediated by p-glycoprotein. *J. Neurochem*. **76,** 1121–1128 (2001).
63. Corder, E. H. *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261,** 921–923 (1993).
64. Tarlungeanu, D. C. *et al.* Impaired amino acid transport at the blood brain barrier is a cause of autism spectrum disorder. *Cell* **167,** 1481–1494 (2016).
65. Dieterich, L. C. *et al.* Transcriptional profiling of human glioblastoma vessels indicates a key role of VEGF-A and TGFβ2 in vascular abnormalization. *J. Pathol.* **228,** 378–390 (2012).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

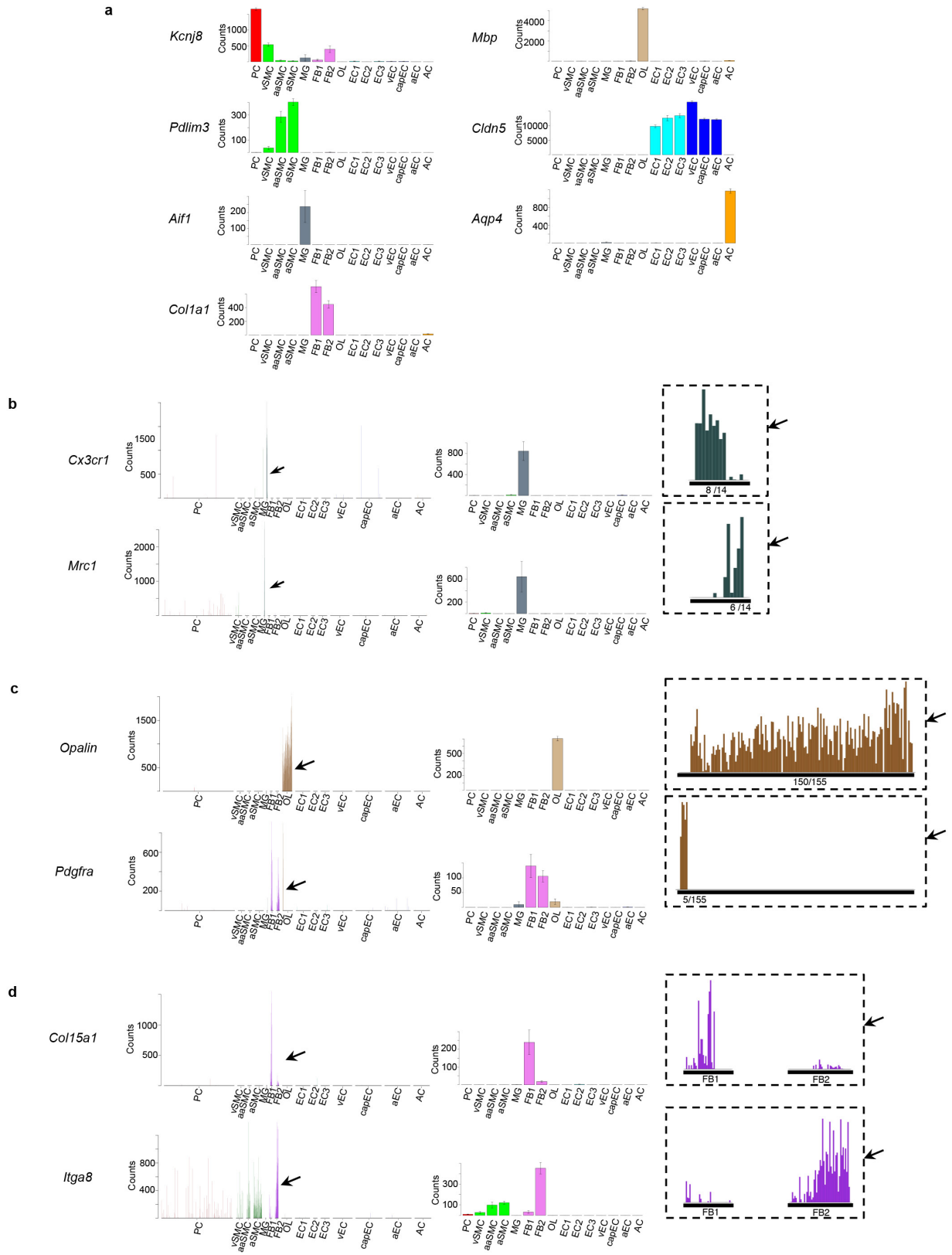


Extended Data Figure 1 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 1 | Cell sorting, clustering and annotation and control for batch effects. a, Representative FACS plots for cell sorts from the various transgenic reporter lines and C57Bl6 controls. For all sorts, the cells were first roughly selected with a broad gate based on forward scatter (FSC)/side scatter (SSC). This initial gate was then used as parent gate for cell selection based upon the relevant fluorescent signal of the different reporters. No additional gating was included, and the cut-off between positive and negative fluorescent signals was assessed using non-fluorescent C57Bl6 mice. **b**, BackSPIN analysis of cells obtained from *Cldn5*-GFP, *Pdgfrb*-GFP;*Cspg4*-DsRed and *Pdgfra*-H2BGFP mice. The BackSPIN algorithm splits the cells stepwise and also assigns genes uniquely to one branch with each split. At split level 1, all endothelial cells were assigned to the right branch, whereas all other cell types were assigned to the left branch. BackSPIN continues to split for as long as it recognizes expression heterogeneity among the genes assigned to that cluster leading to that some clusters continue to split dichotomously, whereas other clusters stop splitting at a certain level. Cluster 0, containing pericytes, stopped splitting at level 2. Any level can be arbitrarily chosen for cluster annotation; we chose level 6, at which 38 clusters had formed. **c**, The expression of the housekeeping gene *Actb* across the 38 clusters in **b** is illustrated as a bar plot, where each bar represents one cell (~3,500 in total). A gross annotation of the 38 clusters based on cell class-specific marker expression (see Fig. 1 for examples) is depicted below the bar plot. **d**, Abbreviations for the cell types. **e**, A reanalysis of cluster 0 from **b**, **c** by BackSPIN using all genes generated dichotomous splits over six levels, showing that this cluster had stopped spitting in **b** because differentially expressed genes had been assigned to other branches. **f**, In the 64 clusters generated in **e**, expression of SMC markers, such as *Acta2,* was found in the right-most clusters. These clusters were subsequently identified as vSMC. **g**, The same tree as in **b**, with cluster 0 split in to PC and vSMC at level 7. Some clusters at level 6 appeared similar and were consolidated to a higher node, as indicated by the red boxes. **h**, The resulting simplified tree, which also shows the final cell type annotation that is used to display the data throughout the paper and at http://betsholtzlab.org/VascularSingleCells/database.html. **i**, An excerpt from the database illustrates the expression of *Actb* across the cell types. Cluster 1 at level 6 in **b**, **g** contained four cells displaying astrocyte markers; this cluster was considered too small and was removed from the dataset. Astrocytes were instead obtained from *Tagln-Cre;R26-stop-tdTomato* mice and added to the dataset for comparison. **j**, The position of the major cell classes in a *t*-SNE diagram, as annotated using marker expression. **k**, Representative marker distribution for the cell classes in *t*-SNE diagrams. **l**, **m**, Control for batch effects. Neither *t*-SNE nor SPIN showed any clustering bias related to mouse individual or experimental repetitions. **l**, The distribution of 1,385 mural cells colour-coded for their mouse origin in *t*-SNE (top) and SPIN diagrams (bottom: an arbitrarily chosen pan-mural cell marker *Myl9* is shown). **m**, The distribution of 1,100 endothelial cells in *t*-SNE (top) and SPIN (bottom: the zonated endothelial-specific transcript *Mfsd2a* is shown).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

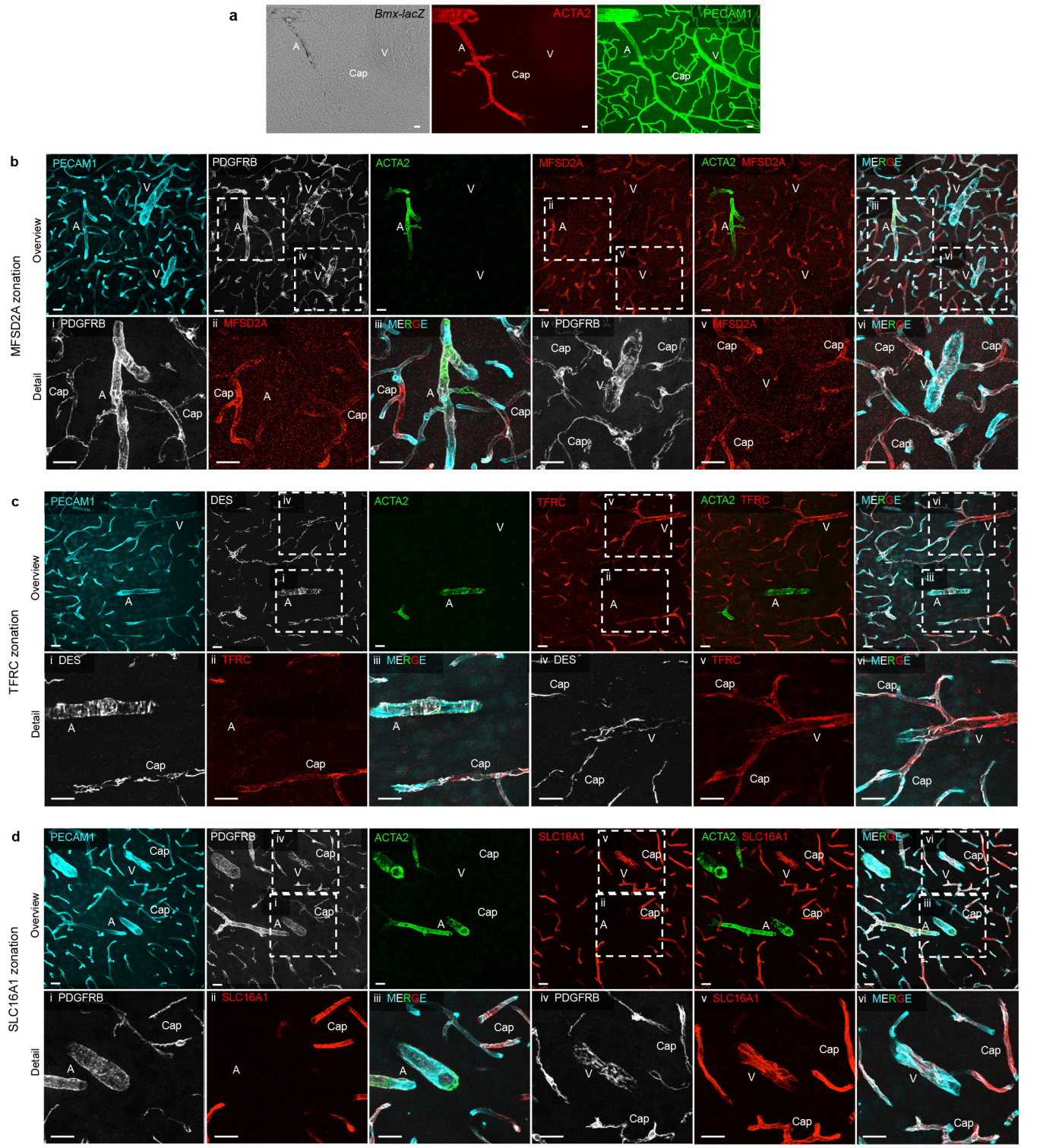


Extended Data Figure 2 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

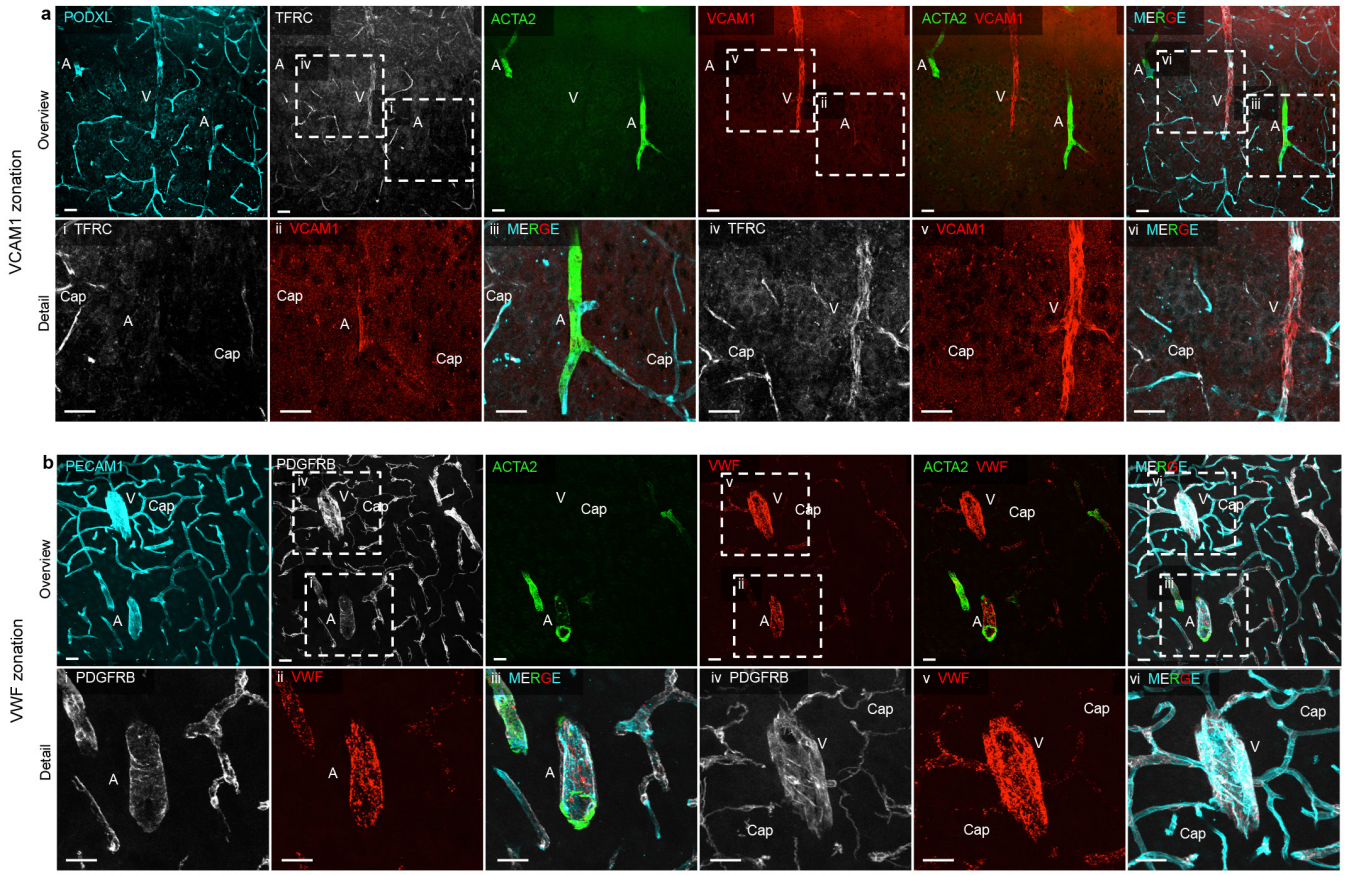
Extended Data Figure 2 | Average expression and cell sub-types. **a**, Average level of expression in the different cell types of the markers shown in Fig. 1b. A key to the cell type abbreviations is provided in Extended Data Fig. 1. Diagrams of the same type for all genes are available at http://betsholtzlab.org/VascularSingleCells/database.html. **b**, Microglial subtypes. The microglial markers *Cx3cr1* and *Mrc1* both mark the microglial (MG) cluster, as shown in the bar-plot diagrams (left) and in the average diagram (middle). Right: insets show magnified views of bar-plots indicated by black arrows. Of the 14 microglial cells present in the MG cluster, the eight left-most cells express *Cx3cr1* but not *Mrc1*, whereas the six right-most cells show the reverse pattern. Since *Cx3cr1* is a marker for interstitial microglia and *Mrc1* is a marker for perivascular microglia and macrophages, this pattern suggests that both these subtypes of microglia are present in the MG cluster, and that they are separated by SPIN to opposite ends of the cluster. **c**, Oligodendrocyte (OL) subtypes. The marker for differentiating oligodendrocytes *Opalin* is expressed in the 150 right-most out of the 155 cells in the OL cluster, whereas the oligodendrocyte precursor cell (OPC) marker *Pdgfra* is expressed only in the five left-most cells (note that *Pdgfra* is also highly expressed in the FB1 and FB2 clusters). Since the OLs were sorted from an initial preparation of brain vascular fragments (in turn isolated using anti-PECAM1 coated magnetic beads) from *Pdgfra*-H2BGFP mice, the *Opalin*-positive cells appear physically associated with blood vessels. As for the microglial subtypes, SPIN separates the two oligodendrocyte lineage cell subtypes to opposite ends of the cluster. **d**, Subtypes of fibroblast-like cells. Numerous markers distinguish cells in the FB1 and FB2 clusters. Here we show the examples *Col15a1*, which is expressed preferentially in FB1 cells, and *Itga8,* which is expressed preferentially in FB2 cells.

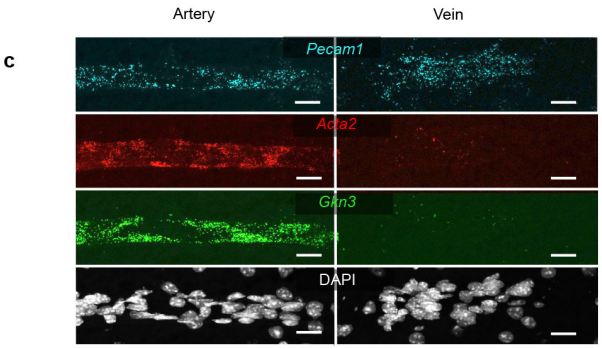
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

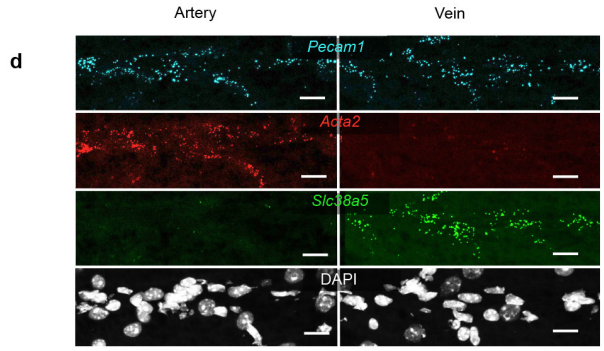


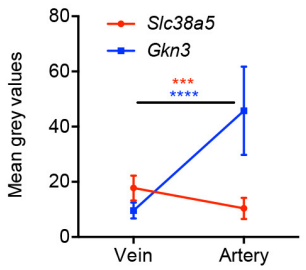
Extended Data Figure 3 | Endothelial arteriovenous zonation of *Bmx*, MFSD2A, TFRC and SLC16A1. All panels show microscopic fields from cerebral cortex with different labels for endothelial cells (cyan: PECAM1), mural cells including SMCs and pericytes (white: PDGFRB or DES), arterial SMCs (green: ACTA2), specific zonated proteins (red: MFSD2A, TFRC or SLC16A1), separately, or in the indicated combinations. A, artery or arteriole; V, vein or venule; Cap, capillary/ies. **a**, Validation of *Bmx* as an arterial marker in adult mouse brain using *Bmx-lacZ* mice. ACTA2 immunofluorescence distinguishes arteries from veins and capillaries. PECAM1 immunofluorescence depicts all vessels. **b**–**d**, The first row of each panel shows the same microscopic field at overview magnification for different immunofluorescence stainings. Hatched boxes delineate the areas shown at higher magnifications in the second row of each panel. The boxes are individually labelled i–vi. Arteries and arterioles are identified by ACTA2. Veins and venules are identified by vessel diameter in combination with the absence of ACTA2. Capillaries are identified by diameter and association with PDGFRB or DES-positive pericytes. An alternative choice of PDGFRB or DES labelling for mural cells was made to match the co-staining with other antibodies. **b**, Zonated expression of MFSD2A. Note the strong expression in capillaries and weak or absent staining in arteries/arterioles and veins. **c**, Zonated expression of TFRC. Note the strong expression in capillaries and veins/venules but absent staining in arteries/arterioles. **d**, Zonated expression of SLC16A1. Note the strong expression in capillaries and veins/venules but absent expression in arteries/arterioles. Data were reproduced in sections taken from three different mice. Scale bars, 20 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.







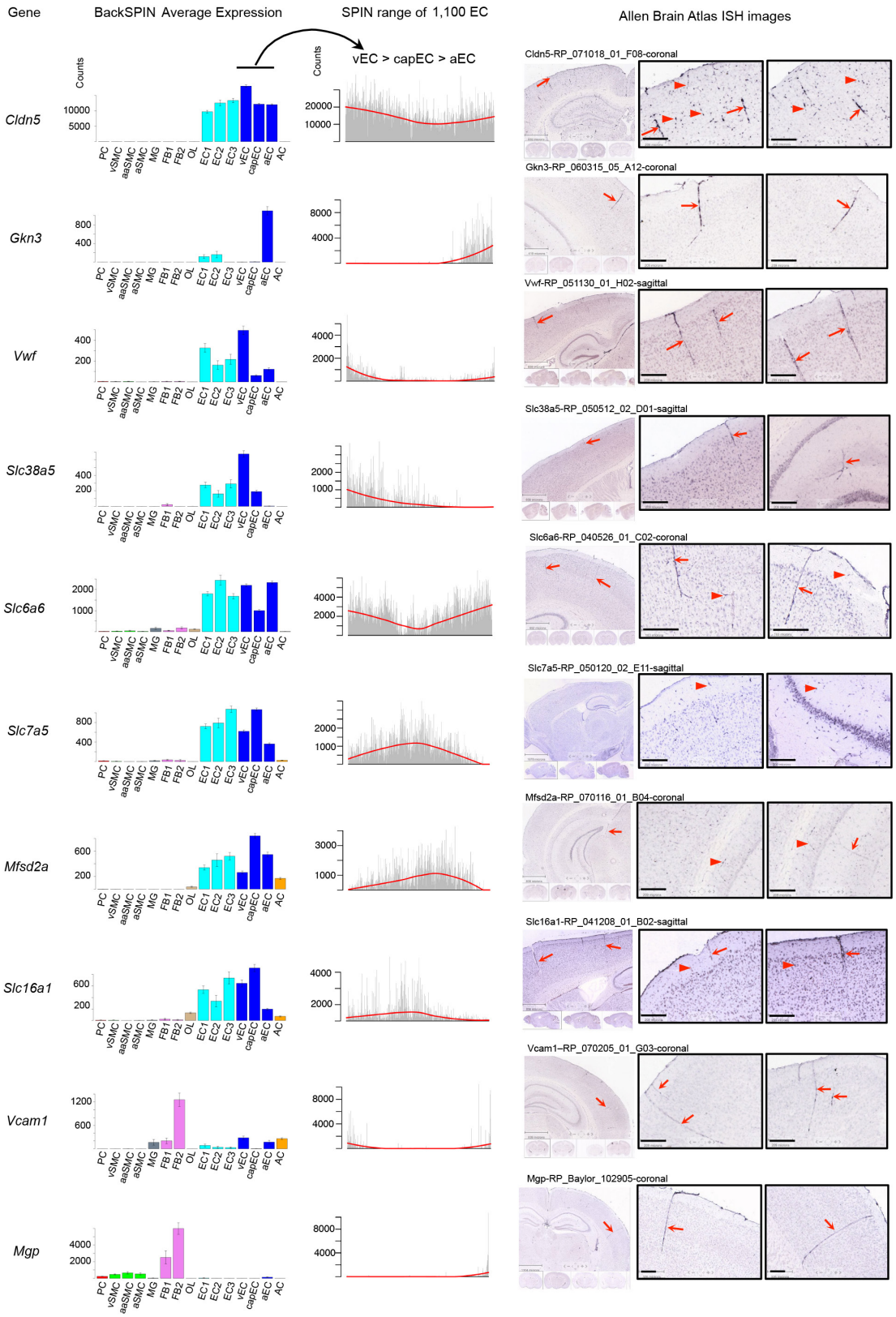


Extended Data Figure 4 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 4 | Endothelial arteriovenous zonation of VCAM1, VWF, *Gkn3* and *Slc38a5*. All panels show data from cerebral cortex. **a**, **b**, Immunofluorescence with different labels for endothelial cells (cyan, PECAM1 or PODXL), mural cells including SMC and pericytes (white, PDGFRB), arterial SMCs (green, ACTA2), and specific zonated proteins (red, VCAM1 or VWF), separately, or in the indicated combinations. The first row of each panel of **a**, **b** shows the same microscopic field at overview magnification for different immunofluorescent stainings. Hatched boxes delineate the areas shown at higher magnifications in the second row of each panel. The boxes are individually labelled i–vi. Arteries and arterioles (A) are identified by ACTA2. Veins and venules are identified by vessel diameter in combination with the absence of ACTA2. Capillaries are identified either by TFRC-positive staining and diameter, or by diameter and association with PDGFRB-positive pericytes. **a**, Zonated expression of VCAM1. Note the strong expression in veins, weak but clear expression in arteries/arterioles, and absent expression in capillaries. **b**, Zonated expression of VWF. Note the strong expression in arteries and veins, weak expression in arterioles and venules and weak expression in capillaries. **c**, **d**, RNA ISH for the indicated genes validates the zonal distribution of *Gkn3* to arteries and *Slc38a5* to veins. **e**, Quantification of the ISH results. The following number of mice and fields in total were analysed: *Slc38a5* ($n = 3$, 15 fields); *Gkn3* ($n = 2$, 20 fields). In Fiji software the ROI manager was used to mark and measure the mean grey values of the vascular segments of interest (vein, artery) and data were presented using GraphPad Prism 6 software. An unpaired *t*-test with Welsh correction was used to demonstrate statistical significance in gene expression between arteries and veins. *x*-axis shows the vascular segment, *y*-axis shows the mean grey values. Unpaired *t*-test with Welsh correction demonstrates statistically significant difference for both genes (****$P < 0.0001$, ***$P < 0.001$). Data were reproduced in sections taken from three different mice. Scale bars, 20 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

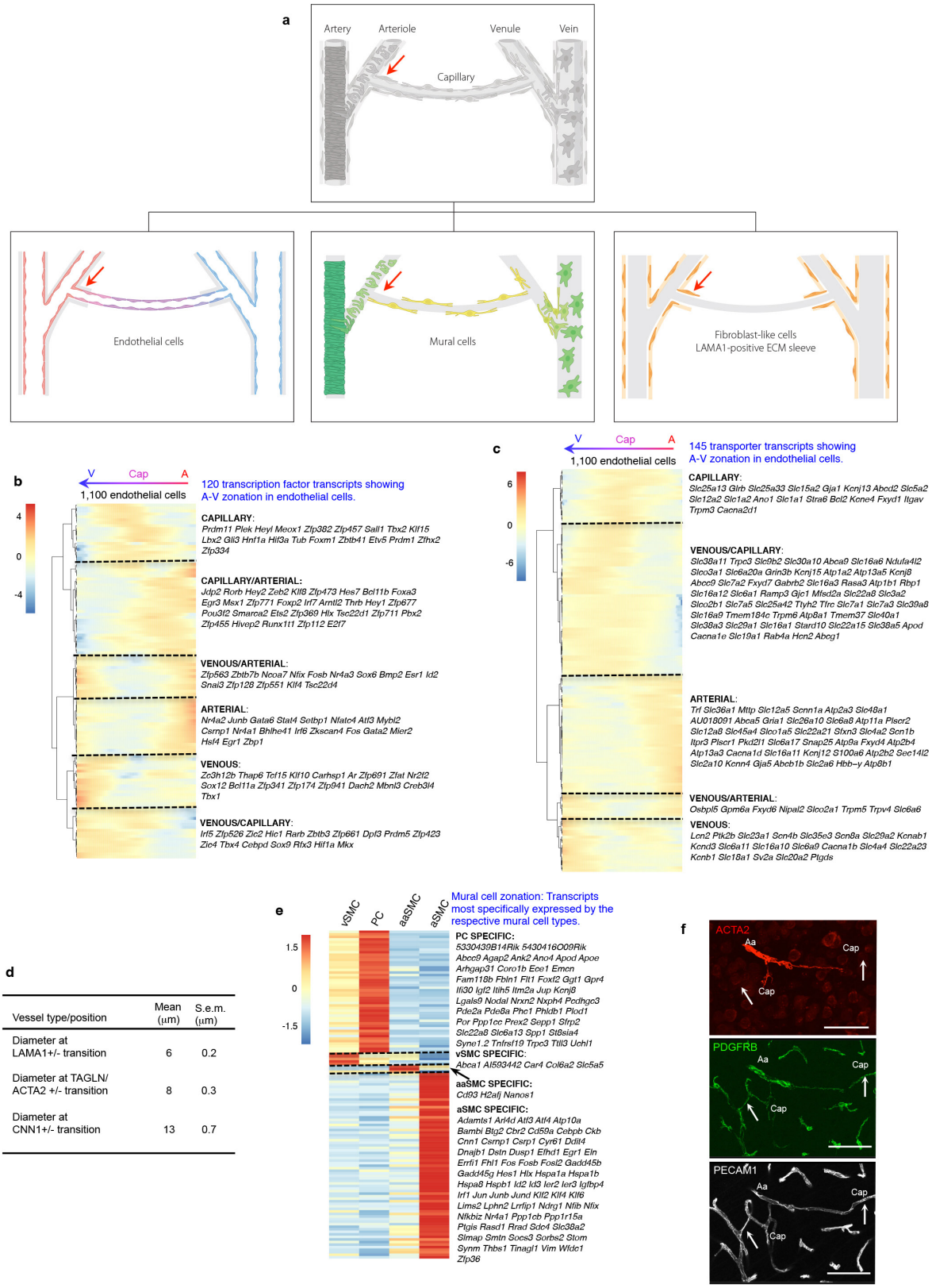


Extended Data Figure 5 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 5 | Validation of endothelial arteriovenous zonated gene expression using Allen Brain Atlas. Allen Brain Atlas data for nine zonated endothelial transcripts (*Gkn3*, *Vwf*, *Slc38a5*, *Slc6a6*, *Slc7a5*, *Mfsd2a*, *Slc16a1*, *Vcam1* and *Mgp*) and one highly expressed endothelial transcript that does not show statistically significant zonation (*Cldn5*). The left panel of diagrams shows the expression distribution across all vascular cell types clustered by BackSPIN and depicted by average expression for each cluster. These diagrams are excerpts from http://betsholtzlab.org/VascularSingleCells/database.html, where bar plots can also be accessed and viewed at high magnification. A key to the cell type abbreviations is provided in Extended Data Fig. 1. The middle panel of diagrams shows bar-plots for 1,100 endothelial cells (the clusters indicated by the black arrow) organized in one dimension by SPIN. The grey bars show the expression level in individual cells and the red curve (loess function) depicts the expression pattern across the range. The arteriovenous zonation indicated above the middle panel of diagrams is suggested by the data shown in Fig. 2 and Extended Data Figs 3, 4: venous endothelial cells to the left, capillary in the middle and arterial cells to the right. The right panel of micrographs shows excerpts from the Allen Brain Atlas, one overview and two at higher magnification (black frames). Genes with preferential expression in arteries (for example, *Gkn3*, *Mgp*), veins (for example, *Slc38a5*) or both (for example, *Vwf*, *Vcam1* and *Slc6a6*) show distinct signals in large penetrating as well as in deep vascular structures. Genes that are highly expressed in capillaries (*Mfsd2a*, *Slc16a1*, and *Cldn5*) show distinct labelling of small vessel structures at all locations. The strong expression of *Vcam1* and *Mgp* in fibroblast-like cells may contribute to the labelling of large vessels, although immunofluorescence shows a preferential localization of *Vcam1* to endothelial cells (Fig. 3). Red arrows point to labelled profiles consistent with arteries/arterioles or veins. Red arrowheads point to profiles consistent with microvessels. Allen Brain Atlas experiment codes are provided above each overview image. Scale bars, 200 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



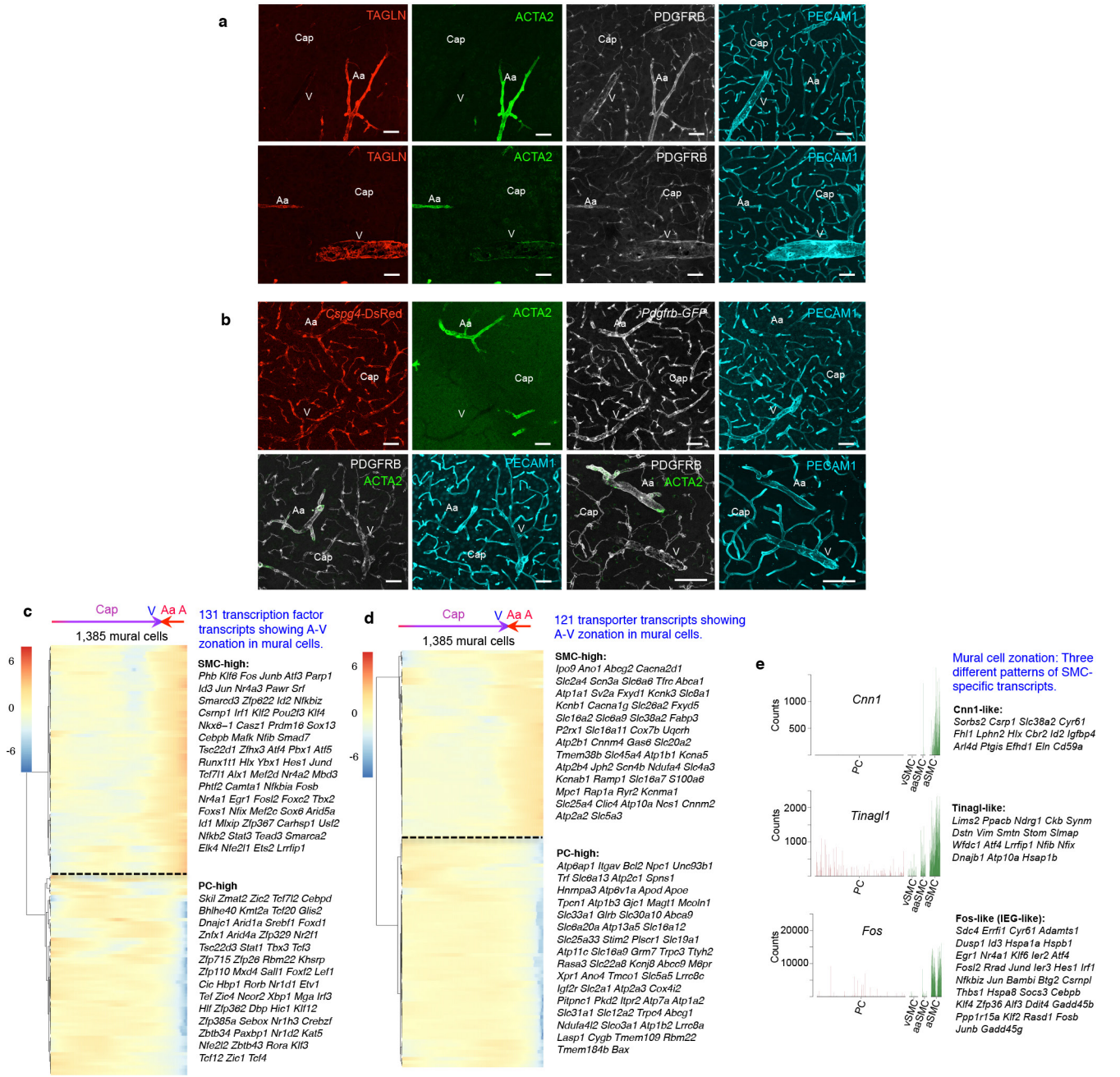
Vessel type/position	Mean (μm)	S.e.m. (μm)
Diameter at LAMA1+/- transition	6	0.2
Diameter at TAGLN/ ACTA2 +/- transition	8	0.3
Diameter at CNN1+/- transition	13	0.7

Extended Data Figure 6 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

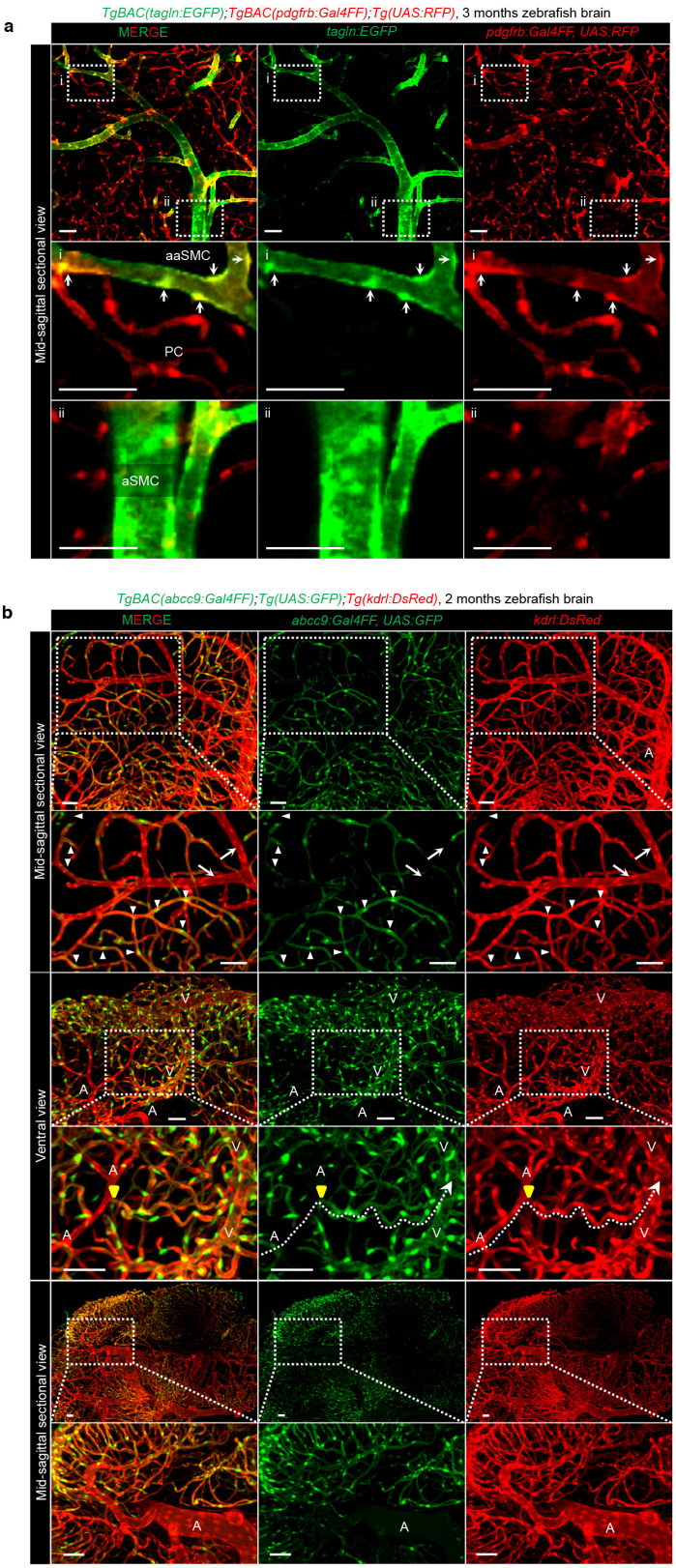
Extended Data Figure 6 | Zonation principles, endothelial transcription factors and transporters, and genes specific to mural cell subtypes. **a**, Schematic illustration of zonation principles and distribution of cell types. Endothelial arteriovenous zonation: red–blue; mural arteriovenous zonation: green–yellow; fibroblast-like cells and LAMA1 sleeve: brown. Arrow, arteriolar-capillary boundary. **b**, **c**, Heat map of relative expression (red, high; blue, low) of 120 transcription factor (**b**) and 145 transporter (**c**) transcripts across the SPIN range of 1,100 endothelial cells. Venous (V), capillary (Cap) and arterial (A) position in the range is depicted at the top. The identities of the transcripts distributing according to six or five major patterns of zonation, respectively, are provided to the right of the heat maps. **d**, Morphometric analysis of vessel diameter at the transitions between cells positive (+) or negative (−) for immunofluorescent labelling of the indicated proteins. The following number of mice and vessel profiles were analysed for each value: diameter at LAMA1 +/− transition, 5 mice, 49 vessels; diameter at TAGLN/ACTA2 +/− transition, 3 mice, 66 vessels; diameter at CNN1+/− transition, 3 mice, 34 vessels. **e**, Heat map of relative expression (red, high; blue, low) of the transcripts most specific for each subtype of mural cells ordered according to anatomical axis (vSMC, PC, aaSMC and aSMC). Transcript identities for each group are provided to the right of the heat map. **f**, Immunofluorescence shows abrupt transition from arteriolar SMCs to pericytes (arrows). Scale bars, 50 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 7 | Zonation of mural cell markers in mouse brain. a, **b**, Immunofluorescence with different labels for endothelial cells (cyan, PECAM1), and mural cells including SMCs and pericytes (white, PDGFRB), arterial SMC (green, ACTA2), specific zonated proteins/reporters (red, TAGLN or CSPG4–DsRed) or in the indicated combinations. All images are from cerebral cortex. These panels show protein distribution to vessel types that match the zonal distribution of transcripts to mural cell subtypes, as shown in Fig. 3a. Data were reproduced in sections taken from three different mice. Scale bars, 50 μm. **c**, **d**, Heat map of relative expression (red, high; blue, low) of 131 transcription factor (**c**) and 121 transporter (**d**) transcripts across the SPIN range of 1,385 mural cells. The anatomical positions (capillary (Cap), venous (V), arteriolar (Aa) and arterial (A)) are shown at the top. Note the presence of two zones and their non-alignment with the anatomical axis (punctuated zonation where arterioles and veins meet). The identity of the transcripts distributing according to the two major patterns of zonation (SMC-high and PC-high) is provided to the right. **e**, Examples of three arterial SMC transcript distribution patterns exemplified by *Cnn1*, *Tinagl1* and *Fos* and other transcripts with similar distribution. The transcripts with *Fos*-like distribution include products of numerous immediate early genes (IEGs), including *Egr1*, *Fosb*, *Jun* and *Junb*.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

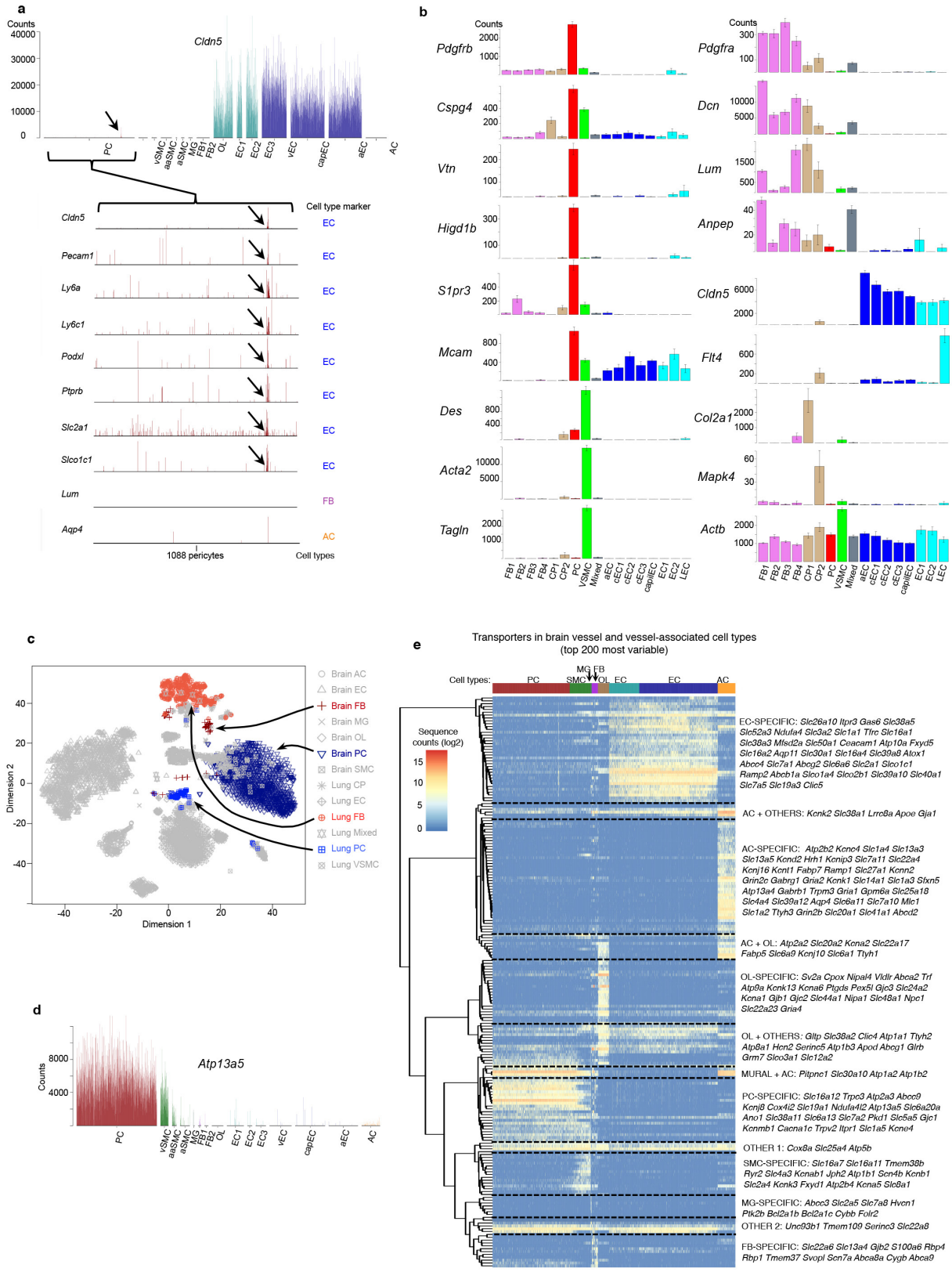


Extended Data Figure 8 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 8 | Zonation of mural cell markers in zebrafish brain. a, The zonated distribution of *tagln* and *pdgfrb* as shown in the three-month-old zebrafish brain using transgenic reporter fish. GFP is expressed from the *tagln* promoter and RFP from the UAS promoter activated by *pdgfrb* promoter-driven Gal4FF. Similar to mice, *tagln* is expressed in a/aaSMCs (white arrows, inset i) whereas *pdgfrb* marks mural cells along all vessel types, including capillary pericytes (PC). The lower panels show insets at high magnification. Note the lower expression of *pdgfrb* in the aSMCs of the larger arteries (inset ii). **b**, GFP expression from a UAS promoter activated by Gal4FF expressed from the *abcc9* promoter in forebrain mural cells of a two-month-old zebrafish. Endothelial cells are labelled by DsRed expression from the *kdrl* (*Vegfr2*) promoter. Note that, similar to the expression pattern of *Abcc9* in mice, fish *abcc9* expression is localized mainly to capillary pericytes and venous mural cells, but not to arteries or arterioles (A). Top, a midsagittal view shows GFP-positive pericytes (white arrowheads) on capillary branches, whereas the SMC of the feeding arteriole as well as the upstream artery (A, white arrows) are GFP-negative. Middle, in a ventral view of the capillary network branching off an arteriole (A), numerous GFP-positive pericytes are seen, while the arteriole is negative. An arteriole–capillary transition is indicated by a yellow arrowhead. Dashed line indicates a vascular continuum stretching from the arteriole (A) across capillary branches to a venule (V). Bottom, another midsagittal view illustrates that arterial SMCs (A) lack *abcc9* expression. Scale bars, 50 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

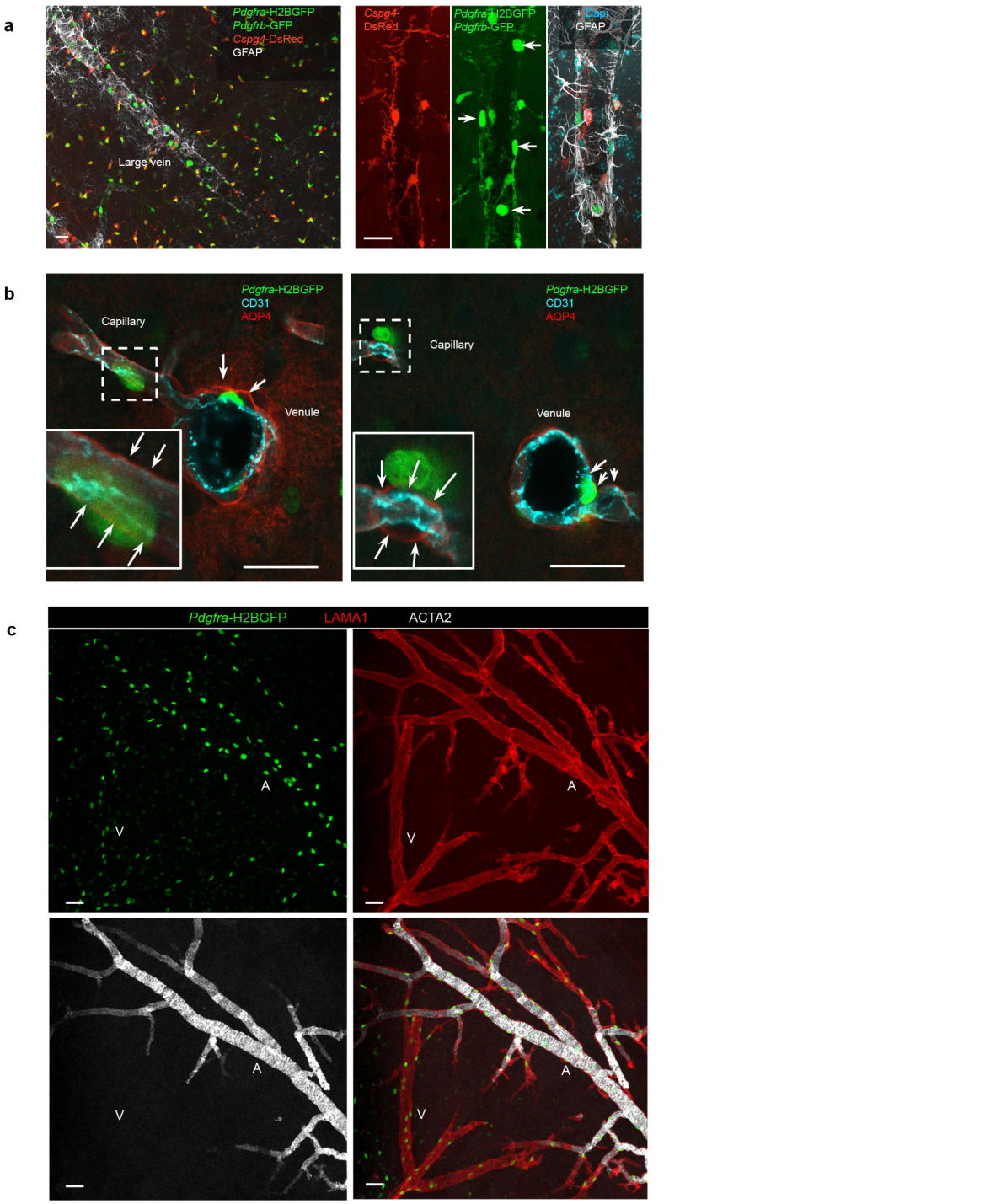


Extended Data Figure 9 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

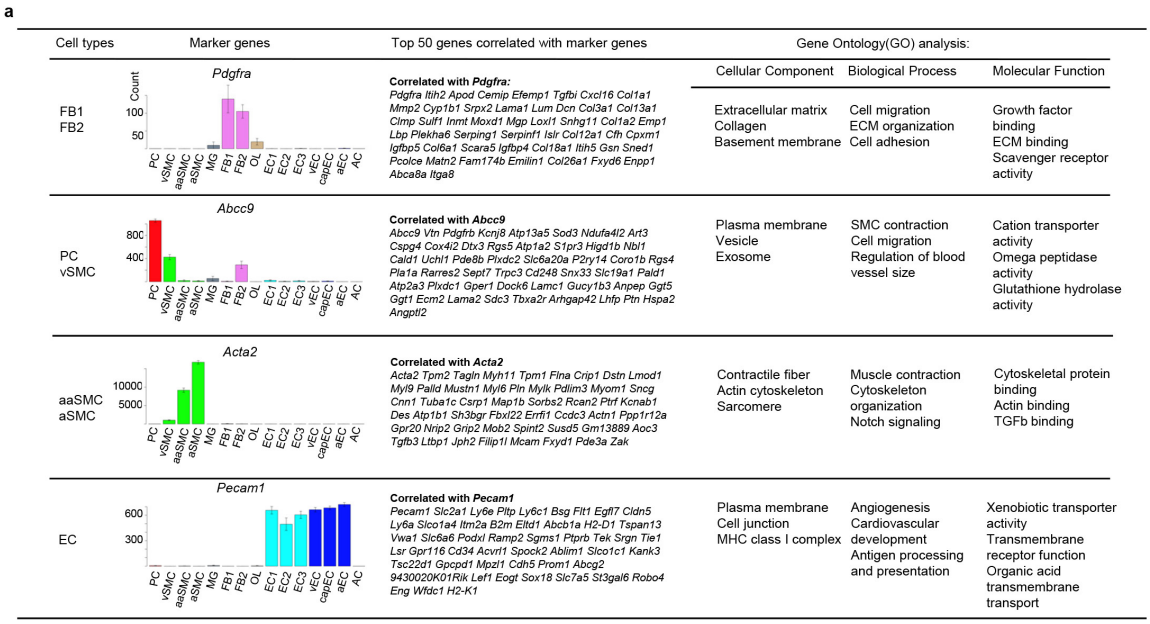
Extended Data Figure 9 | Pericyte definition, organotypicity and transporter transcript expression. a, Bar plot showing the expression of the highly abundant endothelial marker *Cldn5* across the BackSPIN clusters; note the abundant and specific expression in EC clusters. The arrow points to a small group of cells in the PC cluster showing *Cldn5* expression. The BackSPIN range of 1,088 pericytes is shown at larger magnification in the lower panels. An additional seven high abundant endothelial markers (*Pecam1*, *Ly6a*, *Ly6c1*, *Podxl*, *Ptprb*, *Slc2a1* and *Slco1c1*) all show a peak of expression in the same roughly 10 cells, suggesting endothelial contamination of these pericytes. The high abundance markers for fibroblast-like cells (*Lum*) and astrocytes (*Aqp4*) do not reveal signs of contamination of the pericytes by these cell types. **b**, Annotation of lung vascular cell BackSPIN clusters and identification of lung pericytes. The PC cluster was identified according to the expression of known and/or shared markers with brain pericytes, combined with the absence of markers for vascular smooth muscle cells (VSMC), fibroblasts (FB), chondrocytes/perichondrium (CP), or endothelial cells (EC). The mixed cluster contains markers for epithelial cells together with markers for fibroblasts and probably represents contaminated cells. Among the endothelial cells (EC), several clusters were identified, including arterial (a), capillary (c/capil) and lymphatic (LEC). **c**, *t*-SNE analysis of brain and lung cells shows separation of brain and lung pericytes, as well as of brain and lung fibroblast-like cells, into distinct clusters suggesting organotypic differences for both cell classes. **d**, Bar plot of expression of the transporter transcript *Atp13a5* in brain shows that it is highly specific for pericytes and vSMCs. **e**, Heat map illustrating the expression (red, high; blue, low) of the 200 most variable transporter transcripts across the brain vascular and vessel-associated cell types. Hierarchical clustering divides the patterns of transporter distribution into approximately 13 groups (separated by hatched lines). The identity of the transcripts in each group is provided to the right of the heat map. Note the abundance of transporters that are pericyte-specific.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



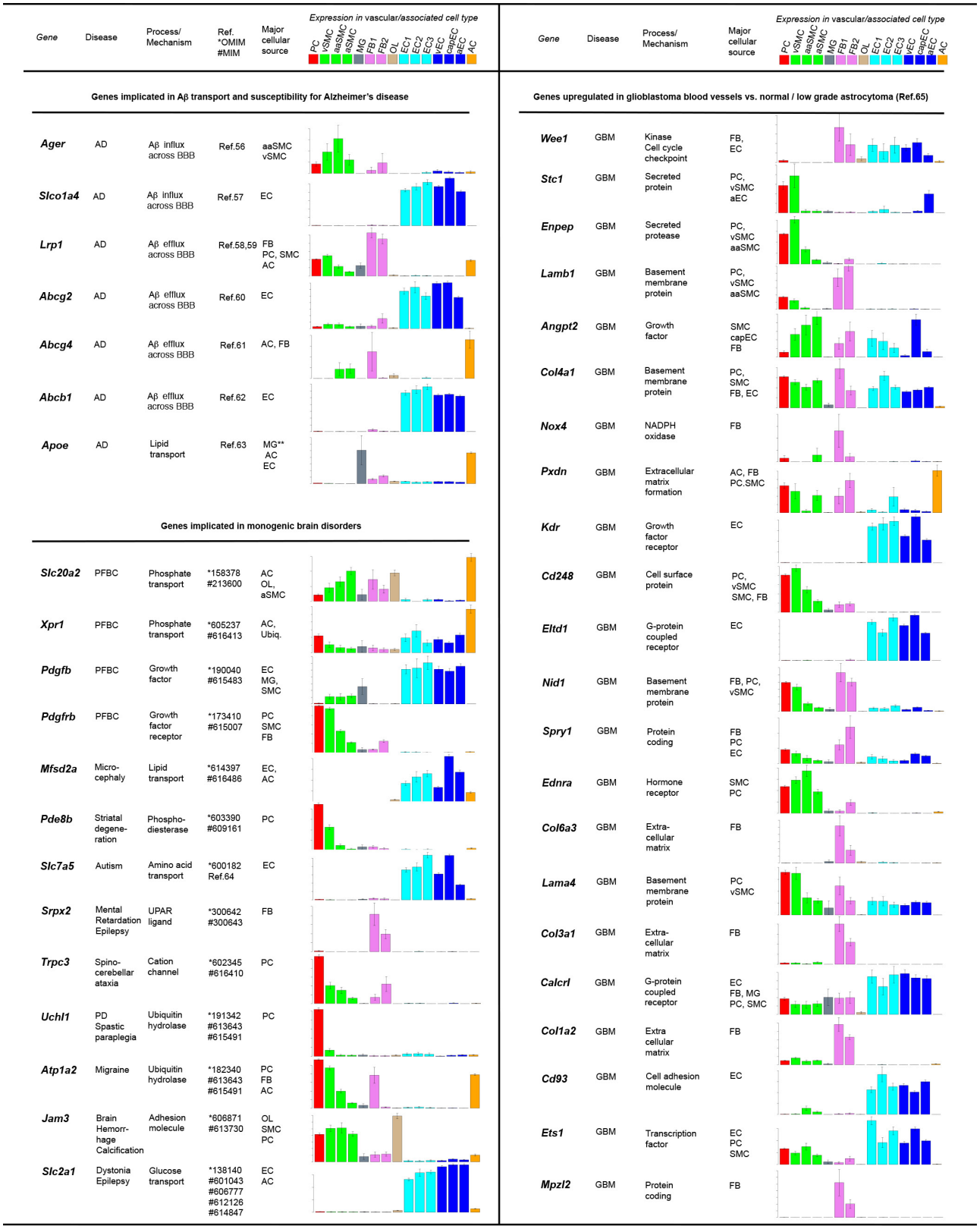
Extended Data Figure 10 | Localization of perivascular fibroblast-like cells. a, *Pdgfra*-H2BGFP-positive cells in veins. Right panels show that the H2BGFP-positive nuclei are distinct from *Cspg4*-DsRed- and *Pdgfrb*-GFP-positive mural cells and GFAP-positive astrocytes. White arrows indicate *Pdgfra*-H2BGFP-positive nuclei. **b**, *Pdgfra*-H2BGFP-positive nuclei associated with venules are located between the endothelial cells and AQP4-positive astrocyte end-feet. Capillary-associated *Pdgfra*-H2BGFP-positive nuclei are located outside the astrocyte end-feet (insets); these cells probably represent oligodendrocyte lineage cells. **c**, The distribution of LAMA1 correlates with the distribution of vessel-associated *Pdgfra*-H2BGFP-positive cells. Image in **c** is from striatum and images in **a** and **b** are from cerebral cortex. Data were reproduced in sections taken from three different mice. Scale bars, 20 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 11 | Characterization of fibroblast-like cells in the brain. a, GO analysis of the 50 most specific transcripts for the following four cell categories: fibroblast-like cells (FB1, FB2), pericytes and venous SMCs (PC, vSMC), arteriolar and arterial SMCs (aaSMC, aSMC) and endothelial cells (EC). The transcripts were selected on the basis of their distribution correlation with *Pdgfra*, *Abcc9*, *Acta2* and *Pecam1*, respectively. Right panel shows a selection of significant terms for each cell class in the three mentioned categories. **b**, Further categorization of the 50 top markers for fibroblast-like cells reveals that most of them are associated with ECM components, turnover, or function. A few are related to innate immunity function. Nearly all markers are also expressed by lung fibroblasts. **c**, **d**, Expression heat map (blue, low; red, high) of collagen transcripts in mural cells and fibroblast-like cells. **c**, Absolute expression. **d**, Relative expression. Note the abundance of collagen transcripts in the fibroblast-like cell clusters (FB1 and FB2) and the difference between them, indicating FB subtypes. **e**, Bar plot distribution of *Lum* shows strong specificity for the FB clusters. Immunofluorescence for LUM shows its specific localization to blood vessels larger than capillaries and a correlation with the distribution of *Pdgfra*–H2BGFP-positive perivascular cells. **f**, Bar plot distribution of *Dcn* shows strong specificity for the FB clusters. Allen Brain Atlas images confirm expression in large penetrating brain vessels (red arrows).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 12 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 12 | Vascular cell type-specific expression of genes linked to brain diseases. Brain vascular scRNA-seq data provide insight into brain disease mechanisms. Three types of disease mechanisms are illustrated. For genes implicated in the transport of Alzheimer's amyloid-β (Aβ) peptide across the BBB, it is noteworthy that several different cell types of the neurovascular unit are implicated in both Aβ influx and efflux mechanisms. Surprisingly, *Lrp1* shows its highest expression in perivascular fibroblast-like cells (FB) and pericytes (PC) or vSMCs, but no detectable expression in endothelial cells (EC). *OMIM and #MIM numbers are provided for the monogenic brain disorders. We list the four genes known to be mutated in primary familial brain calcification (PFBC) to illustrate the different cellular expression of these genes, suggesting that changes in cell–cell communication in the vascular wall plays a role in the pathogenesis of this disease. It is noteworthy that several of the genes mutated in other listed monogenic brain disorders are commonly expressed in the cell types molecularly defined through the present study, such as *Srpx2* (FB) and *Uchl1* (PC). Other transcripts, including *Slc2a1* and *Mfsd2a,* show the expected expression in EC. Unexpectedly, *Jam3*, which is thought to be expressed mainly in EC, has its major vascular expression in mural cells and oligodendrocytes (OL). Of genes overexpressed in glioblastoma grade IV (GBM) vessels, it is striking that most have their normal expression site in FB and PC or vSMC. This may suggest that GBM vasculature has an altered cellular composition, including a higher abundance of mural and perivascular cells, possibly reflecting a fibrotic reaction originating from FB. The diagrams display relative levels of expression. High-resolution quantitative versions of the diagrams and single cell bar plots can be viewed at http://betsholtzlab.org/VascularSingleCells/database.html. **A detailed analysis of the expression of the microglial expression of *Apoe* shows that it is predominantly expressed in perivascular brain macrophages.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ARTICLE

doi:10.1038/nature25453

Structure of the mechanically activated ion channel Piezo1

Kei Saotome[1,2], Swetha E. Murthy[1], Jennifer M. Kefauver[1,2], Tess Whitwam[1], Ardem Patapoutian[1] & Andrew B. Ward[2]

Piezo1 and Piezo2 are mechanically activated ion channels that mediate touch perception, proprioception and vascular development. Piezos are distinct from other ion channels and their structure remains poorly defined, impeding detailed study of their gating and ion permeation properties. Here we report a high-resolution cryo-electron microscopy structure of the mouse Piezo1 trimer. The detergent-solubilized complex adopts a three-blade propeller shape with a curved transmembrane region containing at least 26 transmembrane helices per protomer. The flexible propeller blades can adopt distinct conformations, and consist of a series of four-transmembrane helix bundles we term 'Piezo repeats'. Carboxy-terminal domains line the central ion pore, and the channel is closed by constrictions in the cytosol. A kinked helical beam and anchor domain link the Piezo repeats to the pore, and are poised to control gating allosterically. The structure provides a springboard to dissect further how Piezos are regulated by mechanical force.

The conversion of mechanical force into cellular signals, or mechanotransduction, is an essential biological function conserved throughout evolution. Mechanically activated (MA) channels confer force sensitivity to cells and organisms by allowing the passage of ions across the membrane in response to a mechanical stimulus[1]. Piezo1 and Piezo2 were identified in 2010 as a conserved class of eukaryotic MA ion channels[2,3] involved in various mechanotransduction pathways, including touch sensation[4], proprioception[5], nociception[6], vascular development[7], and breathing[8]. In humans, gain-of-function (GOF) mutations in Piezo1 are linked to dehydrated hereditary stomatocytosis (DHS)[9–12], a disease in which red blood cells are dehydrated and have increased permeability to cations, and GOF Piezo2 mutations are associated with distal arthrogryposis[13]. Loss-of-function (LOF) mutations in Piezo1 and Piezo2 are linked to congenital lymphatic dysplasia[14] and defective touch perception and proprioception[15–17], respectively. Piezos are structurally unrelated to other proteins, and are uniquely large compared to other ion channels (>2000 amino acids). Electrophysiological studies have demonstrated that Piezo1 ion channels are gated directly by changes in membrane tension[18–20], but the structural bases of mechanosensitivity and ion conduction in Piezo channels remain unknown. Previous crystallographic[21,22] and medium-resolution cryo-electron microscopy (cryo-EM)[22] studies have suggested the overall architecture of Piezos. However, a mechanistic understanding of Piezo function requires high-resolution structural information of the channel.

Cryo-EM sample and structure determination

A previous cryo-EM reconstruction of mouse Piezo1[22] lacked amino acid side chain definitions and connectivity of resolved regions, precluding building of an accurate structural model. Here, we optimized Piezo1 sample preparation parameters to make Piezo1 amenable to high-resolution single particle cryo-EM structure determination (Extended Data Fig. 1). After computationally sorting particle heterogeneity (Extended Data Fig. 2a), we identified a subset of particles that were refined into both asymmetric and C3 symmetry-imposed maps of the mouse Piezo1 trimer, with overall resolutions of 4.5 and 3.9 Å, respectively. We conducted masked refinement of the central 'core' (Extended Data Fig. 2b-d) and C3 symmetry expansion followed by masked classification and refinement (Extended Data Fig. 3) to improve densities proximal to the central three-fold axis and the protein periphery, respectively. Both the 3.8 Å resolution core-masked map and a 4.1 Å resolution symmetry-expanded map (Blade Class 1) were used to build and refine a molecular model of mouse Piezo1 that spans structured regions of residues 1134–2544, including 14 TM helices per protomer and other conserved C-terminal domains (Extended Data Fig. 4, Extended Data Table 1, Supplementary Data 1).

Piezo1 core architecture

The homotrimeric Piezo1 complex exhibits a three-blade propeller shape with a extracellular 'cap' domain at the top of the central axis (Fig. 1), consistent with the previous medium-resolution cryo-EM reconstruction[22] (Extended Data Fig 5a-c). The transmembrane region can be approximated by the weak density corresponding to the detergent micelle (Fig. 1a-b), in agreement with the hydrophobicity of the membrane-exposed amino acids in our model (Fig. 1c). Interestingly, the transmembrane region has a pronounced bend, suggesting Piezo1-induced local distortion of the membrane. Though we cannot rule out that the bending is a product of detergent solubilization, protein-induced membrane curvature has also been proposed for mechanosensitive channels of the two-pore potassium[23] and MscS[24,25] families.

The Piezo1 core (Fig. 1d-f) defines an architecture unique from other membrane proteins of known structure. The inner and outer TM helices lie closest to the central axis, and are intervened by the extracellular cap, the structure of which was determined previously by x-ray crystallography[21,22]. In isolation, the trimeric arrangement of inner and outer helices and cap domains bear structural resemblance to acid sensing ion channels[26] and P2X receptors[27], as noted previously[28]. The 12 TM helices peripheral to the inner and outer helices can be grouped into three 4-TM helix-containing bundles each with the same topology that we term 'Piezo Repeats' A, B, C lettered according to proximity to the central axis. Piezo Repeat A of one protomer is adjacent to the outer helix of a different protomer, producing a 'domain-swapped' arrangement analogous to some tetrameric ion channels[29,30] (Fig. 1e). The 'anchor' domain forms a triangular three-helix wedge between the inner/outer helices and Piezo Repeat A. The C-terminal domain (CTD) connects to the intracellular end of the inner helices, collaring the central axis. The intracellular 'beam', is a ~64 residue (~95 Å) kinked helix

[1]Howard Hughes Medical Institute, Department of Neuroscience, The Scripps Research Institute, La Jolla, CA 92037, USA. [2]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California 92037, USA.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ACCELERATED ARTICLE PREVIEW

that runs nearly parallel to the membrane, directly underneath Piezo Repeats A to C. While the N-terminus of each beam is tethered directly to Piezo Repeat C, its C-terminus projects towards the central axis and connects via a ~37 residue disordered linker to a short, coiled peptide we term the 'latch' that contacts the intracellular face of the CTDs. The latch then connects back to Piezo Repeat B via a relatively non-conserved ~240 residue stretch of peptide (Supplementary Data 1) not resolved in our maps. The beam and latch are positioned to transmit conformational changes from the propeller blade periphery to the central ion channel. The beam may also be influenced by a nearby domain that we term the 'clasp' (Fig. 1d,f), which was modeled without amino acid assignment due to lower resolution and discontinuity with other domains.

Flexible blades are composed of at least six 'Piezo Repeats'

Blade-focused classification and refinement of symmetry-expanded particles yielded two classes (Blade Class 1 and Blade Class 2) with improved density at the periphery (Extended Data Fig. 3). Remarkably, in each of these maps, clear density for an additional three Piezo repeats is resolved (Fig. 2a,b,). Thus, a total of six Piezo repeats (which we have named A-F ordered by proximity to the central axis), or 26 TM helices per protomer (including the inner and outer helices) are apparent, in disagreement with a previous cryo-EM study[22], which proposed 14 TM helices. We note the possibility of additional TM helices, and potentially Piezo Repeats, in the Piezo1 protein that were not resolved in our maps.

Although we were only able to confidently place amino acids for Piezo repeats A-C, the topology of the peripheral three repeats is resolved and indicates that all six have the same overall structure and arrangement of TM helices (Fig. 2c,d). The termini of each repeat are cytoplasmic, with a pair of short linkers connecting the N-terminal TM helices, and an extended extracellular linker connecting the third and fourth TM helices. A striking feature of the Piezo repeats is the presence of cytoplasmic membrane-parallel helices tethered to their N-termini. In addition to the Piezo Repeats, the clasp domain also contributes a membrane-parallel helix, adjacent to that of Piezo Repeat B (Fig. 2b). Amino acid assignment for these helices in Piezo repeats A, B, and C demonstrate their amphipathic nature (Extended Data Fig. 6a-d). Interestingly, amphipathic helices are thought to sense or induce membrane curvature in various proteins[31], and have been proposed to be critical for membrane tension-induced gating in various classes of MA channels[32]. That the amphipathic helices dictate local membrane deformation, and perhaps mechanosensitivity of Piezo1 is an intriguing possibility. We speculate the array of amphipathic helices in each blade may work in concert to invoke a local membrane environment with continued curvature, thereby contributing to the puckered architecture.

Our blade-focused classification of the symmetry-expanded particles yielded two classes with distinct conformations demonstrating conformational flexibility of the blade. While the central regions are relatively static, the peripheral regions undergo horizontal movement in the plane of the membrane, not vertically in and out of the membrane (Fig. 2e, Extended Data Fig. 5f,g). Interestingly, associated with this lateral flexibility is a ~5° pivoting of the beam at a kink centered on a highly conserved Ala-Ser-Arg-Gly motif (A1316-G1319)(Fig. 2f, Supplementary Data 1). This pivot may be a critical structural feature that regulates how membrane tension-induced conformational changes at the periphery are transmitted through the beam to the central gating machinery.

Ion conducting pore

The extracellular cap domain, inner helix, and cytosolic CTD line the central pore axis of Piezo1. Extracellular cations could approach the pore entryway vertically through the internal cavity along the three-fold axis of the cap domain, or laterally through gaps between flexible linkers connecting the cap and inner/outer helices (Fig. 3a). The cap domain, which is flexibly linked to the TM domain, might play a role in mediating ion accessibility to the pore, as suggested previously[28]. Indeed, a comparison of our current structure to the previous Piezo1 structure indicates distinct orientations of the cap domain and inner and outer helices (Extended Data Fig. 5d,e), illustrating the flexibility of the cap and the associated TM helices.

Below the cap, the ion conduction pathway is lined by the three inner TM helices. Notably, we find two lateral openings between the inner helices separated by a 'seal' formed by K2479 and F2480 (Fig. 3b). We term these openings the upper crevice (~11 Å wide and ~16 Å tall) and lower fenestration (~10 Å wide and ~12 Å tall), which may provide pore access for lipids or other hydrophobic molecules as a means to regulate ion conduction, reminiscent of the proposed force-gating mechanism of MA two-pore potassium channels[23]. Molecular dynamics simulations of the homotrimeric ATP-gated P2X3 receptor[33] revealed that a lower fenestration provides a passageway for ion flow into the cytoplasm. Though the possibility that the Piezo1 lower fenestration represents an ion permeation pathway warrants further study, we assume that, like most other ion channels, the pore is coincident with the central vertical axis. The pore-facing residues of the inner helices are mainly hydrophobic (Fig. 3d). The ~35 Å long membrane-spanning segment of the pore is relatively wide (> 2 Å van der Waals radius) and does not house the narrowest constrictions of the conduction pathway (Fig. 3c-d). Rather, the ion conduction pathway narrows directly below the lower membrane boundary into a cytosolic ~11 Å long 'neck' (with a minimum van der Waals radius <1 Å) that defines the closed state of the pore in the current structure.

At the top of the cytosolic neck, the side chains of M2493 and F2494, which are conserved across Piezo orthologues (Supplementary Data 1), form a hydrophobic barrier (Fig. 3e). To examine the functional role of the apparent hydrophobic barrier, we recorded MA currents (Fig. 3g,h) from HEK 293T Piezo1 knockout cells[14] expressing the alanine substitution mutants M2493A, F2494A, and M2493A/F2494A. The double mutant M2493A/F2494A resulted in a nonfunctional channel despite proper trimeric assembly and surface expression (Extended Data Fig. 7a-c), demonstrating the importance of these residues. Both single mutants expressed as functional channels, with F2494A having significantly slower inactivation kinetics than wild type for MA macroscopic currents in cell-attached and whole-cell configurations (Fig. 3g,h). Furthermore, stretch activated currents from F2494A had greater steady state currents relative to wild type (Fig. 3i, Extended Data Table 2). The increased channel activity of F2494A in these recordings is consistent with this residue playing a role in gating, possibly by stabilizing a non-conducting state of the channel. The M2493A mutant did not show altered permeation properties, while the F2494A mutant only exhibited modestly increased single channel conductance (Extended Data Fig. 7d,e) and slightly altered ion selectivity (Extended Data Fig. 7f,g, Extended Data Table 2). Our results implicate the hydrophobic barrier as an important contributor to the gating properties of Piezo1, with a minor role in determining ion conduction and selectivity. In previous studies, mutation of acidic residues (E2495 and E2496) adjacent to M2493/F2494 (Fig. 3e) affected pore properties[28], suggesting they might coordinate permeating cations. In the current structure these residues do not project into the pore (Fig. 3e), though structural changes associated with gating could result in this region adopting conformation(s) in which the side chains of E2495 and/or E2496 face the pore to form cation binding site(s).

Directly below the hydrophobic barrier formed by M2493 and F2494, the pore pathway narrows into a second constriction delineated by P2536 and E2537 (Fig. 3f), which are located at the turn of a helix-turn-helix motif in the distal CTD. The side chain of E2537 forms a possible hydrogen bond with Y1412 of the latch domain (Fig. 3f), which may allosterically couple movement of the latch to pore structure in this region. ~4 Å below E2537 a robust density that cannot be reliably modeled but likely originates from Piezo peptide is present (Fig. 3d). This density, which forms a cytosolic 'plug' to the ion channel pore, could also play a role in the permeation pathway.

ACCEPTED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Gating machinery and lipid pocket

Our structure suggests that activation of Piezo1 may involve concerted movements of the inner helix, outer helix, CTD, anchor domain, latch, and beam domains. Interdomain contacts that likely govern gating movements allosterically are shown in Fig. 4. The conserved anchor domain is poised to play a prominent role in gating through its interactions with the inner helix and CTD (Fig. 4b-c). In particular, anchor domain residue E2133 and R2482 of the adjacent inner helix are likely engaged in an electrostatic interaction. In previous studies, mutations at E2133 affected single channel conductance, ion selectivity, and ruthenium red block[34] of Piezo1, and the equivalent mutation to R2482H in human Piezo1 results in slower inactivation kinetics and is associated with DHS[10,12]. In the context of the present structure, these data implicate the anchor domain-inner helix interaction in regulating gating and pore properties of Piezo1. The cytosolic membrane-lining helix of the anchor domain forms several contacts with the CTD (Fig. 4c), including potential electrostatic interaction(s) between R2169 and E2495/E2496, which are adjacent to pore-facing residues M2493 and F2494 (Fig. 3e). This interaction provides a possible explanation for why mutation of E2495 or E2496 affected the ion conduction properties of Piezo1[28]. Overall, the proximity of the anchor domain membrane-lining helix to the permeation pathway's cytosolic 'neck' suggests that its movement would be strongly coupled to channel gating.

We observe a robust lipid-shaped density in an apparent ligand binding pocket formed by the anchor domain on one side and Piezo Repeat A on the other (Fig. 4d). Two conserved arginines (R2035 and R2135) project into the pocket, positioned to stabilize a negatively charged species. Notably, mutation of R2135 to alanine rendered Piezo1 non-functional[34], suggesting that this ligand-binding pocket allosterically supports channel activity, and may therefore be a target for amphiphiles and other compounds aimed at modulating Piezo function. In addition, we note the presence of a separate density potentially representing a lipid headgroup that is sandwiched between Piezo Repeats B and C (Extended Data Fig. 6e-f).

On the cytosolic side, the CTD interacts with the latch and beam (Fig. 4e). Latch residues Y1412, L1414, and F1415 protrude into the surface of the CTD, making contacts with K2528, F2531, E2537, and W2542 (Fig. 4e). Toward the C-terminal end of the beam, a series of basic/polar residues (R1351, R1353, Q1356, and R1360) face acidic CTD residues (D2509, E2518, E2520, and E2524). These electrostatic interactions likely promote allosteric coupling between the beam and CTD. Interestingly, the equivalent mutation to R1353P, which would presumably break or kink the helical beam at this position, is found in human DHS patients and results in delayed inactivation kinetics[10]. Thus, the distal beam end may be a structural determinant of gating behavior, possibly through its interactions with the CTD.

Concluding remarks

The Piezo1 structure described here represents the first high-resolution view of this functionally and structurally novel class of ion channel. Importantly, a majority of known human Piezo1 and Piezo2 GOF disease mutants[35,36] can be mapped onto the structure, which indicates that our model encompasses many of the functionally important domains, and moreover, allows a structural interpretation of disease-causing mechanisms (Fig. 5a).

A central question in the study of Piezos is how membrane tension results in channel activation. A recent study probing for mechanosensitive domains in Piezo1[37] implicated regions as disparate as the N-terminus (which likely resides toward the outer "tip" of the propeller blade) and the central extracellular cap. Thus, mechanical activation is likely to involve movements of various domains spanning the entirety of the protein complex. Our structure clarifies the features that contribute to this process (Fig. 5b-c). The arrayed Piezo Repeats potentially induce a remarkable degree of local membrane deformation, conceivably through their amphipathic helices. Though free energy calculations[38] have suggested that the energetic cost of local membrane deformation plays a key role in the mechanosensitivity of MscS[25] and TRAAK[23] channels, its contribution to Piezo structure and function in membrane bilayers will require further study. Nonetheless, our results allow us to speculate that membrane tension-induced conformational changes of the Piezo Repeats are transmitted through the beam and/or anchor domains to the inner helix and CTD, which in turn determine the activation state of the channel. While the nature of these conformational transitions remain unclear, the present structure of a closed state will guide further investigations of the process of Piezo mechanical activation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 September; accepted 14 December 2017.
Published online 20 December 2017.

1. Ranade, S.S., Syeda, R. & Patapoutian, A. Mechanically Activated Ion Channels. *Neuron* **87**, 1162–79 (2015).
2. Coste, B. *et al.* Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science* **330**, 55–60 (2010).
3. Coste, B. *et al.* Piezo proteins are pore-forming subunits of mechanically activated channels. *Nature* **483**, 176–81 (2012).
4. Ranade, S.S. *et al.* Piezo2 is the major transducer of mechanical forces for touch sensation in mice. *Nature* **516**, 121–U330 (2014).
5. Woo, S.H. *et al.* Piezo2 is the principal mechanotransduction channel for proprioception. *Nature Neuroscience* **18**, 1756–1762 (2015).
6. Kim, S.E., Coste, B., Chadha, A., Cook, B. & Patapoutian, A. The role of Drosophila Piezo in mechanical nociception. *Nature* **483**, 209–212 (2012).
7. Li, J. *et al.* Piezo1 integration of vascular architecture with physiological force. *Nature* **515**, 279–282 (2014).
8. Nonomura, K. *et al.* Piezo2 senses airway stretch and mediates lung inflation-induced apnoea. *Nature* **541**, 176+ (2017).
9. Zarychanski, R. *et al.* Mutations in the mechanotransduction protein PIEZO1 are associated with hereditary xerocytosis. *Blood* **120**, 1908–15 (2012).
10. Albuisson, J. *et al.* Dehydrated hereditary stomatocytosis linked to gain-of-function mutations in mechanically activated PIEZO1 ion channels (vol 4, pg 1884, 2013). *Nature Communications* **4** (2013).
11. Andolfo, I. *et al.* Multiple clinical forms of dehydrated hereditary stomatocytosis arise from mutations in PIEZO1. *Blood* **121**, 3925–3935 (2013).
12. Bae, C., Gnanasambandam, R., Nicolai, C., Sachs, F. & Gottlieb, P.A. Xerocytosis is caused by mutations that alter the kinetics of the mechanosensitive channel PIEZO1. *Proc Natl Acad Sci U S A* **110**, E1162–8 (2013).
13. Coste, B. *et al.* Gain-of-function mutations in the mechanically activated ion channel PIEZO2 cause a subtype of Distal Arthrogryposis. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 4667–4672 (2013).
14. Lukacs, V. *et al.* Impaired PIEZO1 function in patients with a novel autosomal recessive congenital lymphatic dysplasia. *Nat Commun* **6**, 8329 (2015).
15. Chesler, A.T. *et al.* The role of PIEZO2 in human mechanosensation. *New England Journal of Medicine* **375**, 1355–1364 (2016).
16. Mahmud, A.A. *et al.* Loss of the proprioception and touch sensation channel PIEZO2 in siblings with a progressive form of contractures. *Clinical Genetics* **91**, 470–475 (2017).
17. Haliloglu, G. *et al.* Recessive PIEZO2 stop mutation causes distal arthrogryposis with distal muscle weakness, scoliosis and proprioception defects. *Journal of Human Genetics* **62**, 497–501 (2017).
18. Syeda, R. *et al.* Piezo1 channels are inherently mechanosensitive. *Cell Rep* **17**, 1739–1746 (2016).
19. Lewis, A.H. & Grandl, J. Mechanical sensitivity of Piezo1 ion channels can be tuned by cellular membrane tension. *Elife* **4** (2015).
20. Cox, C.D. *et al.* Removal of the mechanoprotective influence of the cytoskeleton reveals PIEZO1 is gated by bilayer tension. *Nature Communications* **7**(2016).
21. Kamajaya, A., Kaiser, J.T., Lee, J., Reid, M. & Rees, D.C. The structure of a conserved piezo channel domain reveals a topologically distinct beta sandwich fold. *Structure* **22**, 1520–7 (2014).
22. Ge, J. *et al.* Architecture of the mammalian mechanosensitive Piezo1 channel. *Nature* **527**, 64–9 (2015).
23. Brohawn, S.G., Campbell, E.B. & MacKinnon, R. Physical mechanism for gating and mechanosensitivity of the human TRAAK K+ channel. *Nature* **516**, 126–30 (2014).
24. Pliotas, C. *et al.* The role of lipids in mechanosensation. *Nature Structural & Molecular Biology* **22**, 991–998 (2015).
25. Phillips, R., Ursell, T., Wiggins, P. & Sens, P. Emerging roles for lipids in shaping membrane-protein function. *Nature* **459**, 379–385 (2009).
26. Jasti, J., Furukawa, H., Gonzales, E.B. & Gouaux, E. Structure of acid-sensing ion channel 1 at 1.9 A resolution and low pH. *Nature* **449**, 316–23 (2007).
27. Kawate, T., Michel, J.C., Birdsong, W.T. & Gouaux, E. Crystal structure of the ATP-gated P2X(4) ion channel in the closed state. *Nature* **460**, 592–8 (2009).

ACCELERATED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

28. Zhao, Q. *et al.* Ion permeation and mechanotransduction mechanisms of mechanosensitive piezo channels. *Neuron* **89**, 1248–63 (2016).
29. Long, S.B., Campbell, E.B. & Mackinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K+ channel. *Science* **309**, 897–903 (2005).
30. Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107–12 (2013).
31. Drin, G. & Antonny, B. Amphipathic helices and membrane curvature. *FEBS Lett* **584**, 1840–7 (2010).
32. Bavi, N. *et al.* The role of MscL amphipathic N terminus indicates a blueprint for bilayer-mediated gating of mechanosensitive channels. *Nat Commun* **7**, 11984 (2016).
33. Mansoor, S.E. *et al.* X-ray structures define human P2X(3) receptor gating cycle and antagonist action. *Nature* **538**, 66–71 (2016).
34. Coste, B. *et al.* Piezo1 ion channel pore properties are dictated by C-terminal region. *Nature Communications* **6**(2015).
35. Wu, J., Lewis, A.H. & Grandl, J. Touch, tension, and transduction - the function and regulation of piezo ion channels. *Trends Biochem Sci* **42**, 57–71 (2017).
36. Glogowska, E. *et al.* Novel mechanisms of PIEZO1 dysfunction in hereditary xerocytosis. *Blood* (2017).
37. Wu, J., Goyal, R. & Grandl, J. Localized force application reveals mechanically sensitive domains of Piezo1. *Nat Commun* **7**, 12939 (2016).
38. Ursell, T., Kondev, J., Reeves, D., Wiggins, P. & Phillips, R. The role of lipid bilayer mechanics in mechanosensation. in *Mechanosensitive Ion Channels,* Vol. 1 (eds. Kamkin, A. & Kiseleva, I.) 37–70 (Springer, 2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Turner and G. Ozorowski for assistance with EM data collection. We acknowledge early efforts to characterize Piezo1 using EM by E. Wilson-Kubalek and R. Milligan, and S Kakuda for screening purification conditions. We thank A. Sobolevsky for critical reading of the manuscript, and members of the Ward and Patapoutian laboratories for discussion. This work was supported by a Ray Thomas Edwards Foundation grant to A.B.W. and NIH grants NS083174 and DE022358 to A.P. Computational analyses were performed using shared instrumentation funded by NIH 1-S10OD021634. A.P. is an investigator of the Howard Hughes Medical Institute. This is manuscript #29582 from The Scripps Research Institute.

Author Contributions All authors designed experiments and analyzed data. K.S. prepared EM samples, collected and processed EM data, built the structural model, and conducted FSEC analysis. S.E.M. carried out electrophysiological experiments and data analysis. J.M.K. developed sample preparation protocols. T.W. carried out immunostaining experiments. All authors wrote and edited manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.P. (ardem@scripps.edu) and A.B.W. (andrew@scripps.edu).

ACCELERATED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

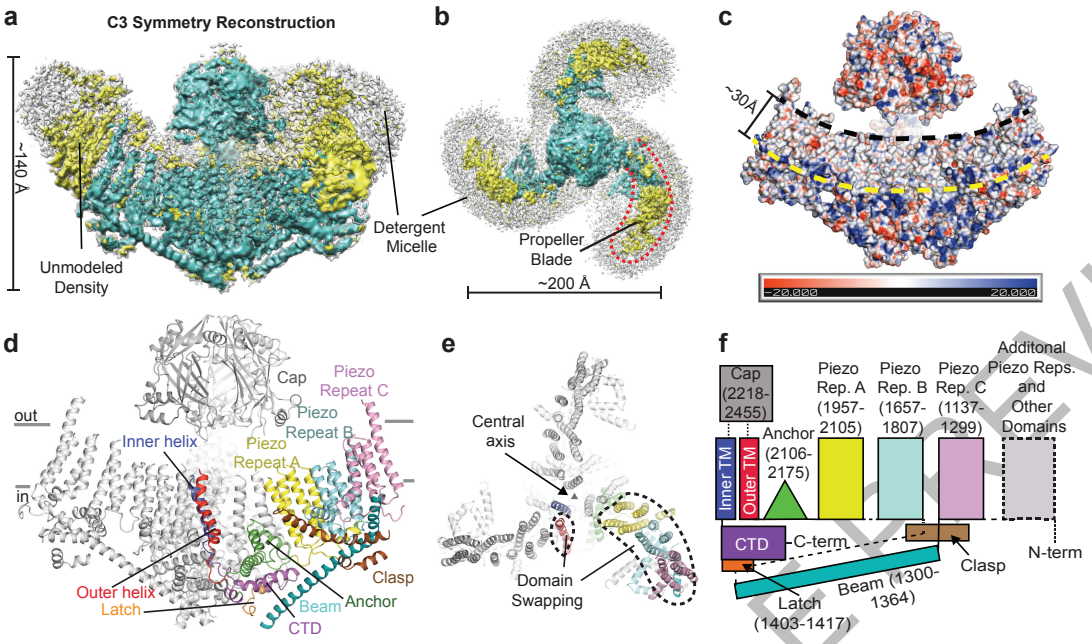


Figure 1 | Architecture and domain arrangement of the Piezo1 core. Side (**a**) and top (**b**) views of the Piezo1 cryo-EM map refined with C3 symmetry imposed. Density corresponding to modeled regions is colored green. Less resolved density that could not be modeled is colored yellow. At lower thresholds (transparent gray map), scattered density likely originating from detergent micelle can be observed. **c**, surface electrostatics of the Piezo1 core model. Note the bent hydrophobic stripe representing the transmembrane region. Dotted lines approximate the extracellular (black) and cytosolic (yellow) membrane boundaries. **d**, **e**, side (**d**) and top (**e**) views of the Piezo1 core in cartoon representation, with each domain colored differently. In **e**, the cap domain is removed to highlight domain swapping between Piezo Repeat A and outer helix. **f**, schematic of the Piezo1 core domain arrangement. Dotted lines represent flexible regions that were not clearly resolved in the density maps.

ACCELERATED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

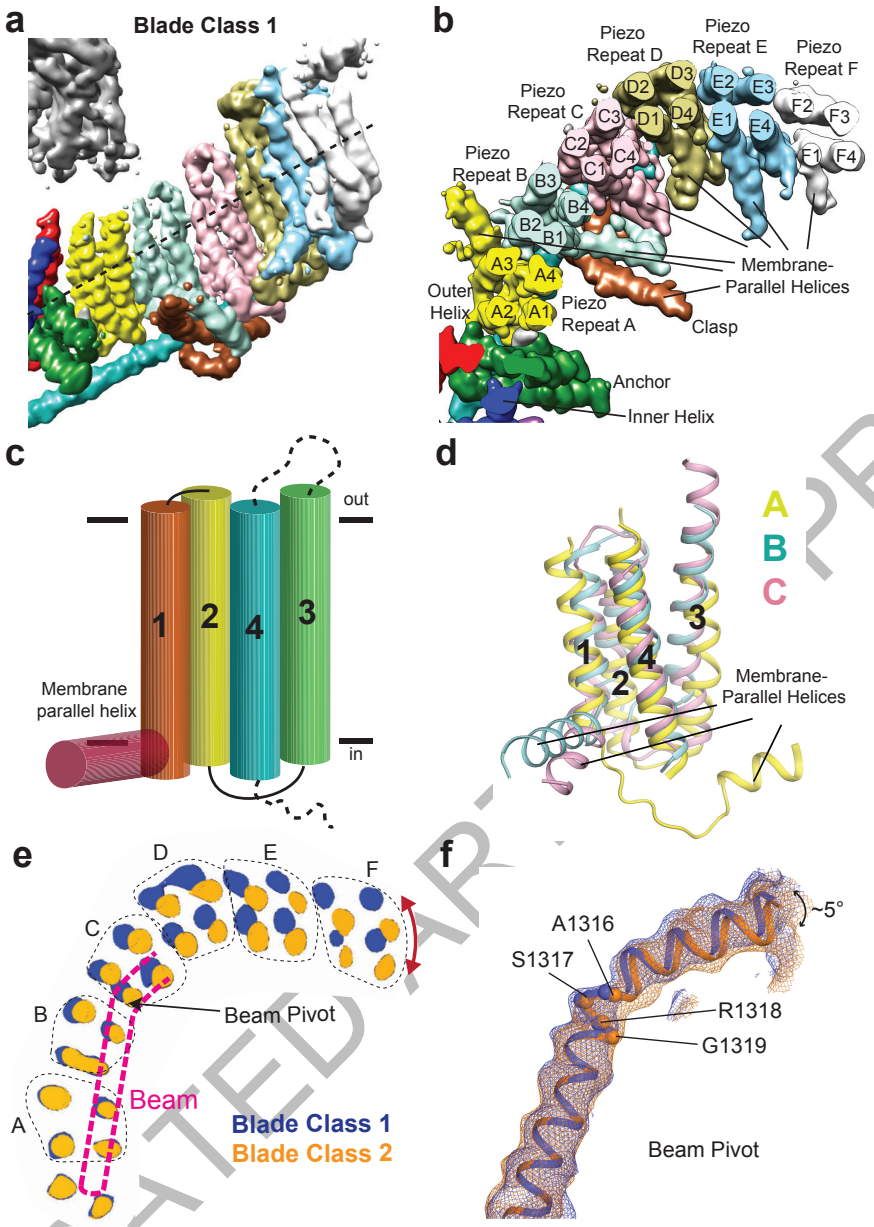


Figure 2 | Propeller blade composition and conformational heterogeneity. a, **b**, side (**a**) and top (**b**) views of Blade Class1 EM map, with individual domains segmented and colored uniquely. In **b,** the map is sliced along the dotted line in **a** to highlight the positions of the TM helices. **c**, schematic of Piezo repeat topology, which includes a cytosolic membrane parallel helix and four TM helices. **d**, Aligned models for Piezo Repeats A (yellow), B (cyan), and C (pink). **e**, Sagittal slice of EM density corresponding to the transmembrane region of Blade Class 1 (blue) and Blade Class 2 (orange). Approximate trajectory of the beam is shown as pink dotted line. Individual Piezo Repeats are circled by black dotted lines and labeled A-F. Red double arrow depicts direction of conformational change between the two classes. **f**, Superposed EM maps and cartoon models corresponding to beam domains of Blade Class 1 (blue) and Blade Class 2 (orange). Cα positions of the conserved Ala-Ser-Arg-Gly motif at the beam pivot are shown as spheres.

ACCELERATED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

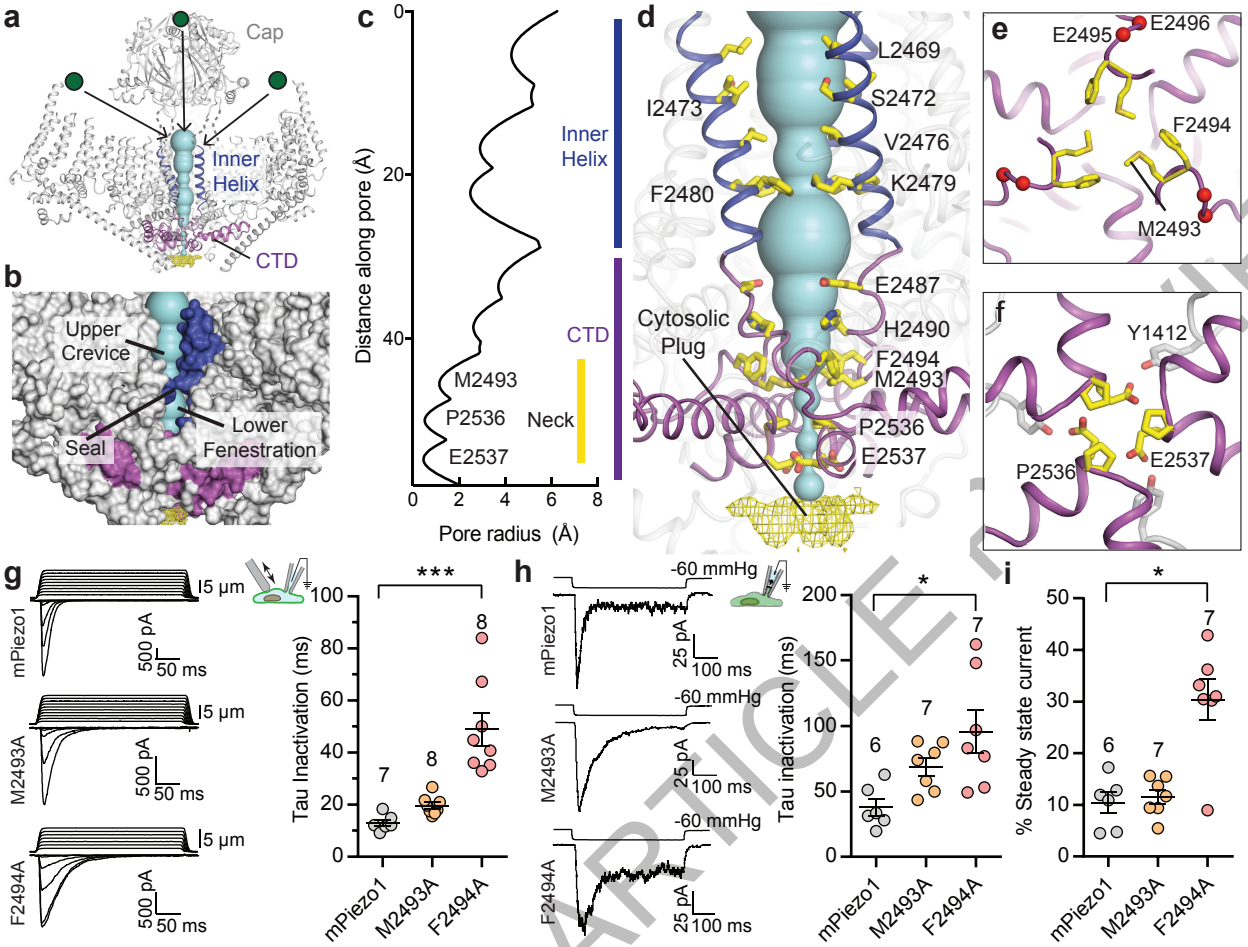


Figure 3 | Ion pore structure and electrophysiological characterization of M2493A and F2494A. a, cartoon model of Piezo1, with inner helix and CTD colored blue and purple, respectively. Flexible linkers between the cap domain and the transmembrane region are depicted as dotted lines. Green spheres represent ions, with possible entryways into the pore shown as black arrows. **b**, surface representation of transmembrane region of Piezo1 surrounding the central axis, highlighting lateral membrane openings to the pore pathway. **c**, van der Waals radii of the pore, plotted against distance along the pore axis. The radial distance between the pore axis and the protein surface is shown as cyan surface in **a**, **b,** and **d. d**, expanded view of the Piezo1 pore, with residues lining the pore pathway shown in yellow and labeled. EM density corresponding to a 'cytosolic plug' is shown as yellow mesh. **e**, **f**, expanded top-down views of constrictions formed by M2493 and F2494 (**e**) and P2536 and E2537 (**f**). Residues mentioned in the text are also labeled. **g**, Left, representative traces of probe displacement and MA whole-cell currents recorded (-80 mV) from one cell expressing wild type, M2493A or F2494A mPiezo1 channels. Right, inactivation time constant for individual cells across different conditions (mPiezo1 N = 7 cells, M2493A N = 8 cells and F2494A N = 8 cells). **h,** Left, representative traces of stretch activated macroscopic currents recorded (-80 mV) in response to -60 mmHg pressure from one cell expressing wild type, M2493A or F2494A mPiezo1 channels. Right, inactivation time constant (at -60 mmHg) for individual cells across different conditions. **i,** Percent steady state current (at -60 mmHg) for individual cells across different conditions. **h**-**i,** mPiezo1 N = 7 cells, M2493A N = 8 cells and F2494A N = 8 cells. **g**-**i**, whiskers represent mean ± s.e.m.; One-way analysis of variance (ANOVA) with Dunn's multiple comparison relative to mPiezo1. **g** ***$P = 0.0001$, **h** *$P = 0.008$ and **i** *$P = 0.012$).

ACCELERATED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

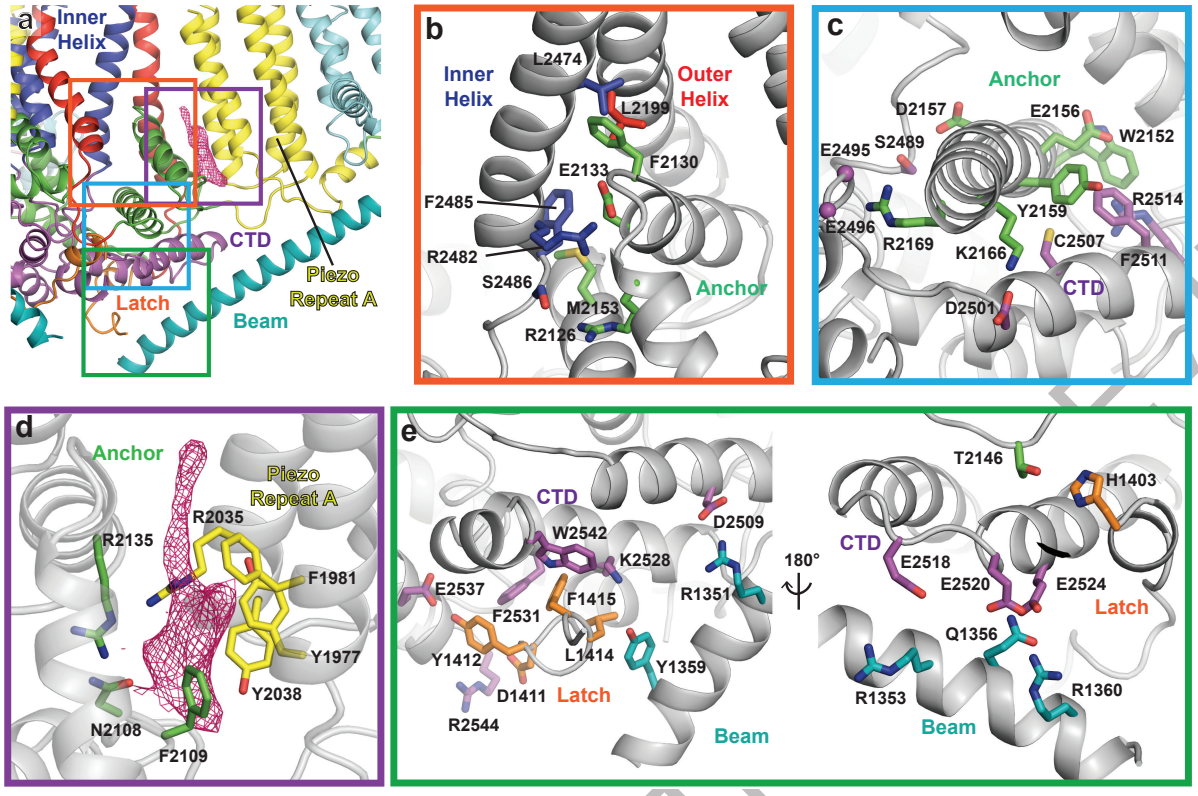


Figure 4 | Interdomain interactions and lipid pocket. a shows cartoon model of Piezo1, with color-coded boxes around regions expanded in **b**-**e**. In **a** and **d**, red mesh shows EM density corresponding to putative lipid. In **a**-**e**, residues contributing to interdomain contacts (**b**, **c**, **e**) or putative lipid binding (**d**) are shown as sticks and colored according to domain.

ACCELERATED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

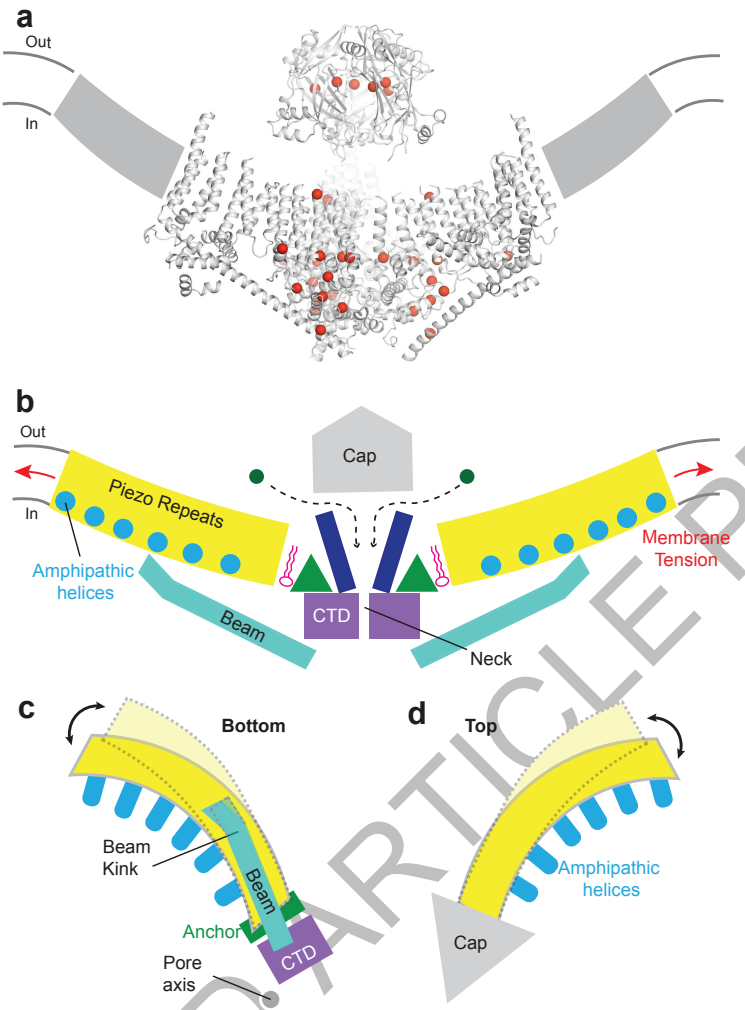


Figure 5 | Mapping of disease mutants and schematic diagram of Piezo1 structure and conformational flexibility. a, Model of the Piezo1 core, with Cα positions of Piezo1 and Piezo2 human GOF disease mutants shown as red spheres. **b-d,** schematic diagrams of the various Piezo1 structural elements that contribute to its function as a membrane-tension gated channel. In **c** and **d**, bottom-up (**c**) and top-down (**d**) views of a single Piezo1 propeller blade are shown, with dotted lines and arrows depicting lateral conformational flexibility, which may be involved in gating.

ACCELERATED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Sample Preparation. Protocols to purify mouse Piezo1 were adapted from previous study[22] with modifications. Suspension-adapted HEK 293F cells (purchased and authenticated from ATCC, tested negative for mycoplasma) were grown at 37 °C to a cell density of ~1.5 million/mL and transfected with a pcDNA3.1-IRES-GFP plasmid[22] containing mouse Piezo1 (Uniprot KB accession number E2JF22) fused to Prescission protease cleavage site (LEVLFQGP) followed by GST coding sequence (this plasmid was a kind gift from Bailong Xiao). 300 μg of DNA and 900 μg of polyethylenimine were used per liter of transfected HEK cells. 4 mM sodium butyrate was added to the cells 12–24 hours after transfection. ~48 hours after transfection, cells were harvested and resuspended in buffer containing 25 mM PIPES pH 7.2, 150 mM NaCl, 1% CHAPS, 0.5% soy phosphatidylcholine, 0.1% $C_{12}E_9$, 2 mM DTT, and cOmplete protease inhibitor cocktail (Roche). The lysate was stirred for 1 hour at 4 C, clarified by centrifugation, and incubated with Glutathione Sepharose 4B resin (GE Healthcare) for 3 hours. The resin was washed extensively with buffer A (25 mM HEPES pH 8.0, 150 mM NaCl, 100 μM glycodiosgenin, 2 mM DTT). Prescission protease was added to a ~30% slurry of resin in buffer A at a ratio of ~100 μg per 1 mL resin and incubated at 4C overnight. The flowthrough was concentrated and subjected to size exclusion chromatography on a Superose 6 increase column equilibrated to buffer A. The peak corresponding to Piezo1 trimer was collected and concentrated to ~5 mg/mL. 3.5 μL of protein was applied to a freshly plasma cleaned UltrAuFoil R1.2/1.3 grid, blotted for 3.5 seconds with a blot force of 0, and plunge-frozen in liquid ethane using a Vitrobot Mark IV (FEI).

Cryo-EM Data collection. An initial data set of mouse Piezo1 was in a Talos Arctica electron microscope (FEI) equipped with a K2 direct detector camera (Gatan) operated at 200 kV and having a nominal pixel size of 1.15 Å. Subsequent data were collected on a Titan Krios electron microscope (FEI) equipped with a K2 direct detector camera (Gatan) operated at 300 kV and having a nominal pixel size of 1.03 Å pixel size. Micrographs were automatically collected using Leginon software[39]. The total accumulated dose was ~60 $e^-/Å^2$ and the target defocus range was 0.8–2.2 μM. Micrograph movies frames were dose-weighted and aligned using MotionCor2[40]. CTF values were estimated using Gctf[41].

Image Processing. Template-based particle picking, 2D and 3D classification, and 3D refinements were carried out in Relion 2.0[42]. Unwanted particle picks due to ice contaminations or the grid hole edges were removed using EM Hole Punch[43]. An initial set of 527 micrographs was collected on the Talos Arctica microscope, from which ~96,000 particle images were extracted. After reference-free 2D classification and 3D classification using the published medium-resolution Piezo1 reconstruction[22] as an initial reference (low pass filtered to 60 Å), a ~26,000 particle subset was identified that refined to a ~5 Å resolution reconstruction with C3 symmetry applied.

Two separate data collection sessions were conducted on the Titan Krios microscope. 1978 micrographs were collected in session 1, yielding ~281,000 particles. This particle stack was subjected to 2D classification to remove classes not displaying well-defined features of the Piezo1 trimer, resulting in ~266,000 particles. In session 2, 1086 micrographs were collected, yielding ~144,000 particles from which ~112,000 particles were chosen after 2D classification. The post-2D classification particle stacks from sessions 1 and 2 were combined into a stack with ~378,000 particles. This stack was subjected to 3D classification requesting 3 classes, using a 60 Å low pass-filtered map of mouse Piezo1 generated from the Talos Arctica data set as an initial reference. The best class contained ~106,000 particles, which were subjected to a round of 3D classification without alignment requesting six classes. The most populated class from this round of refinement contained ~73,000 particles. Auto-refinement of these particles with and without C3 symmetry imposed resulted in reconstructions at 3.9 and 4.5 Å resolution, respectively. C3 symmetry imposed refinement with a soft mask around the Piezo1 core (cap, inner and outer helices, anchor, CTD, Piezo Repeats A-C, latch, and beam domains) requesting local searches resulted in a 3.8 Å resolution map with improved features.

Inspection of the asymmetrically refined map indicated a progressive break of symmetry toward the blade periphery (Extended Data Fig. 3a). To better resolve the peripheral regions of the blade, relion_particle_symmetry_expand was used to artificially enlarge the particle stack three-fold, and add 120° or 240° to the first Euler angle determined in the initial C1 refinement. Alignment-free 3D classification of these stack of ~220,000 "expanded" particles requesting six classes with a soft mask around a single blade yielded two highly populated classes containing ~102,000 (Blade Class 1) and ~74,000 expanded particles (Blade Class 2) with distinct blade conformations. Refinement with a soft mask around a single blade and the most symmetric parts of the molecule (cap, anchor, Piezo Repeat A, inner and outer helices) and requesting local angular searches resulted in reconstructions at 4.1 (Blade Class 1) and 4.5 Å (Blade Class 2) resolution, respectively.

Model Building and Refinement. The 3.8 Å resolution map of the Piezo1 core sharpened with a B factor of -118 automatically determined using relion_post-process was primarily used for model building and refinement. Initially, the crystal structure of the extracellular cap domain[22] (PDB ID: 4RAX) was docked as a rigid body, and a poly-alanine model was fit into the rest of the Piezo1 core. Rough sequence assignment was ascertained from the position of the docked cap domain, as well as transmembrane topology prediction and previous immunostaining and mass spectrometry data[34] demonstrating the positions of extracellular loops and intracellular phosphorylation sites, respectively. We were able to model amino acids for 14 of the 26 TM helices observed in our EM maps. We note here that the previous membrane topology analysis[34] approximates the presence of 21 TM helices in addition to the 14 we have modeled. The N- and C- termini of the channel are intracellular[34], necessitating an even number of TM segments. This would suggest a putative total of 34 or 36 TM helices, and up to 8 Piezo repeats per protomer. Nonetheless, additional studies are needed to clarify the exact membrane topology of Piezo proteins.

Precise sequence registry assignment was determined by locating bulky Phe, Arg, Tyr, and Trp side chains, as well as the absence of side chain density for Gly. Side chain densities were not well resolved for many Asp and Glu side chains, likely due to radiation damage[44]. Some portions of the beam and Piezo Repeats B and C were built using the Blade Class 1 map (sharpened to an automatically determined B-factor of -152), then symmetrized three-fold to fit into the C3 symmetry core map. Iterative manual model building and automatic refinement were carried out in Coot[45], and Phenix.real_space_refine[46], respectively. Final structural refinement was carried out in the C3 symmetry core map, with strong restraints applied to portions built using the Blade Class 1 map, as well as non-crystallographic symmetry restraints for the entire model. Model geometry and fit to map were validated using MolProbity[47] and EMRinger[48]. Electrostatic potential surfaces were calculated using APBS[49]. Structural figures were made in PyMOL[50] or UCSF Chimera[51].

Generation of mutants. mPiezo1(M2493A), mPiezo1(F2494A) and mPiezo1(M2493A, F2494A) mutants were generated using QuickChange II XL site-directed mutagenesis kit according to the manufacturer's instruction and confirmed by full-length DNA sequencing. All mPiezo1 constructs were inserted into an IRES-GFP pcDNA3.1 vector.

Cell culture and transfections. Wild-type and mutant mouse Piezo1 were transfected and tested in Piezo1 deficient Human embryonic kidney 293T (HEK293T P1KO) cells. HEK293T P1KO cells were generated using CRISPR/Cas9 nuclease genome editing technique as described previously[14] and not tested for mycoplasma contamination nor further authenticated. Cells were grown in Dulbecco's Modified Eagle Medium containing 4.5 mg ml^{-1} glucose, 10% fetal bovine serum, 50 units ml^{-1} penicillin and 50 μg ml^{-1} streptomycin. Cells were plated onto 12-mm round glass poly-D-lysine coated coverslips placed in 24-well plates and transfected using lipofectamine 2000 (Invitrogen) according to the manufacturer's instruction. All plasmids were transfected at a concentration of 600 ng ml^{-1}. Cells were recorded from 24 to 36 hours post transfection.

Electrophysiology. Patch-clamp experiments were performed in standard whole-cell or cell-attached mode using Axopatch 200B amplifier (Axon Instruments). Macroscopic currents were filtered at 2 kHz and sampled at 20 kHz. Leak currents before mechanical stimulations were subtracted off-line from the current traces. Voltages were not corrected for a liquid junction potential except for ion selectivity experiments. Liquid junction potential was calculated using Clampex 10.3 software. All experiments were done at room temperature.

Mechanical stimulation: For whole-cell recordings, mechanical stimulation was achieved using a fire-polished glass pipette (tip diameter 3–4 μm) positioned at an angle of 80° relative to the cell being recorded. Downward displacement of the probe towards the cell was driven by Clampex-controlled piezoelectric crystal microstage (E625 LVPZT Controller/Amplifier; Physik Instrumente). The probe had a velocity of 1 μm ms^{-1} during the ramp phase of the command for forward movement and the stimulus was applied for 150 or 300 ms. To assess the mechanical sensitivity of a cell, a series of mechanical steps in 0.5 or 1 μm increments was applied every 20 s. For I–V relationship recordings, voltage steps were applied 700 ms before mechanical stimulation (150 ms) from a holding potential of −60 mV. Voltage steps were given from −80 mV to +80 in 20 mV increments.

Stretch-activated currents were recorded in the cell-attached patch-clamp configuration. Membrane patches were stimulated with a 5 s prepulse duration of +5 or 0 mmHg followed by 500 ms negative pressure pulses through the recording electrode using Clampex-controlled pressure clamp HSPC-1 device (ALA-scientific)[19]. Since the single-channel amplitude is independent of the pressure intensity, the most optimal pressure stimulation was used to elicit responses that allowed single-channel amplitude measurements. These stimulation values were largely dependent on the number of channels in a given patch of the recording cell. Single-channel amplitude at a given potential was measured from trace histograms

ACCELERATED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

of 5 to 10 repeated recordings. Histograms were fitted with Gaussian equations using Clampfit 10.3 software. Single-channel slope conductance for each individual cell was calculated from linear regression curve fit to single-channel I–V plots.

Solutions: For whole-cell patch-clamp recordings, recording electrodes had a resistance of 2–3 MΩ when filled with internal solution composed of (in mM) 133 CsCl, 1 $CaCl_2$, 1 $MgCl_2$, 5 EGTA, 10 HEPES (pH 7.3 with CsOH), 4 MgATP and 0.4 Na_2GTP. The extracellular solution was composed of (in mM) 133 NaCl, 3 KCl, 2.5 $CaCl_2$, 1 $MgCl_2$, 10 HEPES (pH 7.3 with NaOH) and 10 glucose. For ion-selectivity experiments, internal solution used was (in mM) 150 CsCl, 10 HEPES (pH 7.3 with CsOH) and extracellular solution consisted of (in mM) 100 $CaCl_2$ and 10 HEPES (pH 7.3 with CsOH).

For cell-attached patch-clamp recordings, external solution used to zero the membrane potential consisted of (in mM) 140 KCl, 1 $MgCl_2$, 10 glucose and 10 HEPES (pH 7.3 with KOH). Recording pipettes were of 1–3 MΩ resistance when filled with standard solution composed of (in mM) 130 mM NaCl, 5 KCl, 1 $CaCl_2$, 1 $MgCl_2$, 10 TEA-Cl and 10 HEPES (pH 7.3 with NaOH).

Surface immunostaining. Surface immunostaining of mPiezo1 was carried out as described previously[34], with slight modifications. Briefly, pcDNA3.1 IRES-GFP vector encoding wildtype or mutant mPiezo1 constructs containing myc-tag (EQKLISEEDL) inserted at a previously determined extracellular amino acid position 897[34] were expressed in HEK293T P1KO cells on poly-D-lysine treated glass coverslips. 2 days after transfection, labeling of non-permeabilized cells was carried out by incubating the cells with 9E11 anti-Myc antibody (1:50; Santa Cruz Biotechnology). After six washes with warm medium, cells were incubated with secondary antibodies conjugated to Alexa Fluor 568 (1:200; Life Technologies) for 10 min at room temperature. Cells were washed six times with warm media and once with PBS, and then fixed with 4% PFA/PBS for 30 min. For permeabilization, cells were fixed with 4% PFA for 10 min and then treated with 0.3% Triton X-100 and blocked with 10% normal goat serum in PBS prior to incubation with antibodies (primary: 1:200 for 2 hr, secondary: 1:400 for 1 hr, in block solution). Cells were imaged with a Nikon C2 confocal microscope with 40x oil immersion objective by illumination with the 543 nm laser. The live labeling and permeabilized staining were repeated in three separate experiments to confirm results.

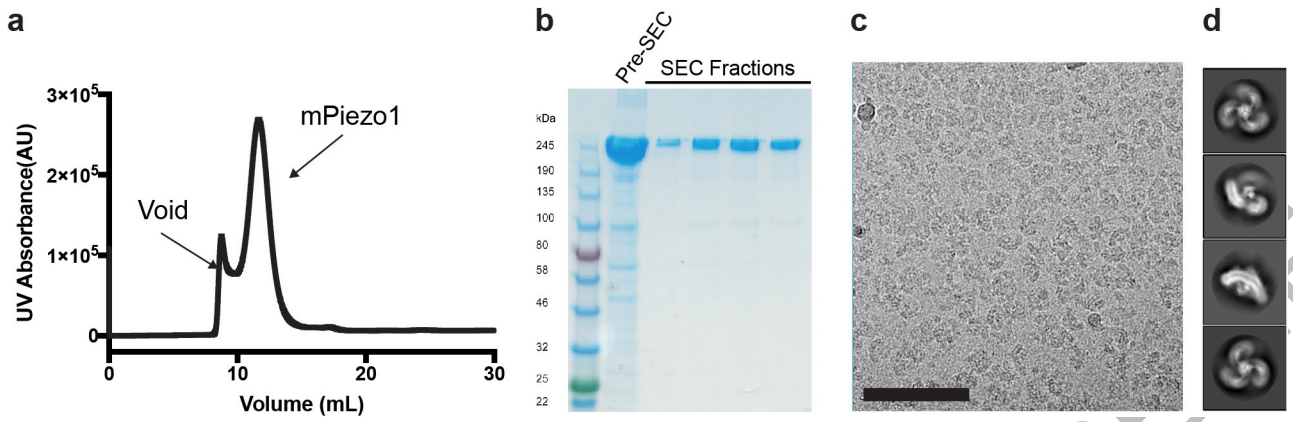
Fluorescence Detection Size Exclusion Chromatography (FSEC). Suspension-adapted HEK293F cells were transfected with pcDNA3.1 vector containing mPiezo1 or mPiezo1-M2493A F2494A double mutant fused to tdTomato at the C-terminus. 48 hours post-transfection, 1 mL of the culture was pelleted and resuspended in 200 μL of buffer containing 25 mM HEPES pH 8.0, 150 mM NaCl, 1% CHAPS, 0.5% soy phosphatidylcholine, 1 mM PMSF, 2 μM pepstatin A, 2 μg/mL aprotinin, 2 μg/mL leupeptin. 100 μL of clarified supernatant was injected into a Superose 6 increase column in line with a fluorimeter tracking tdTomato fluorescence (excitation at 554 nM, emission at 581 nM).

Data Availability. EM Maps were deposited into the Electron Microscopy Data Bank (EMDB) under accession ID 7128. Structural model was deposited to Protein Data Bank (PDB) under accession ID 6BPZ. All other data can be obtained upon request to the corresponding authors.

39. Suloway, C. *et al.* Automated molecular microscopy: the new leginon system. *J Struct Biol* **151**, 41–60 (2005).
40. Zheng, S.Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).
41. Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1–12 (2016).
42. Scheres, S.H. RELION: implementation of a bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519–30 (2012).
43. Berndsen, Z., Bowman, C., Jang, H. & Ward, A.B. EMHP: An accurate automated hole masking algorithm for single-particle cryo-EM image processing. *Bioinformatics* (2017).
44. Bartesaghi, A., Matthies, D., Banerjee, S., Merk, A. & Subramaniam, S. Structure of beta-galactosidase at 3.2-angstrom resolution obtained by cryo-electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 11709–11714 (2014).
45. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
46. Adams, P.D. *et al.* PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* **66**, 213–221 (2010).
47. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D-Biological Crystallography* **66**, 12–21 (2010).
48. Barad, B.A. *et al.* EMRinger: side chain directed model and map validation for 3D cryo-electron microscopy. *Nature Methods* **12**, 943–946 (2015).
49. Unni, S. *et al.* Web Servers and services for electrostatics calculations with APBS and PDB2PQR. *Journal of Computational Chemistry* **32**, 1488–1491 (2011).
50. The PyMOL Molecular Graphics System, Version 1.8 Schrodinger, LLC. (2015).
51. Pettersen, E.F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–12 (2004).

ACCELERATED ARTICLE PREVIEW

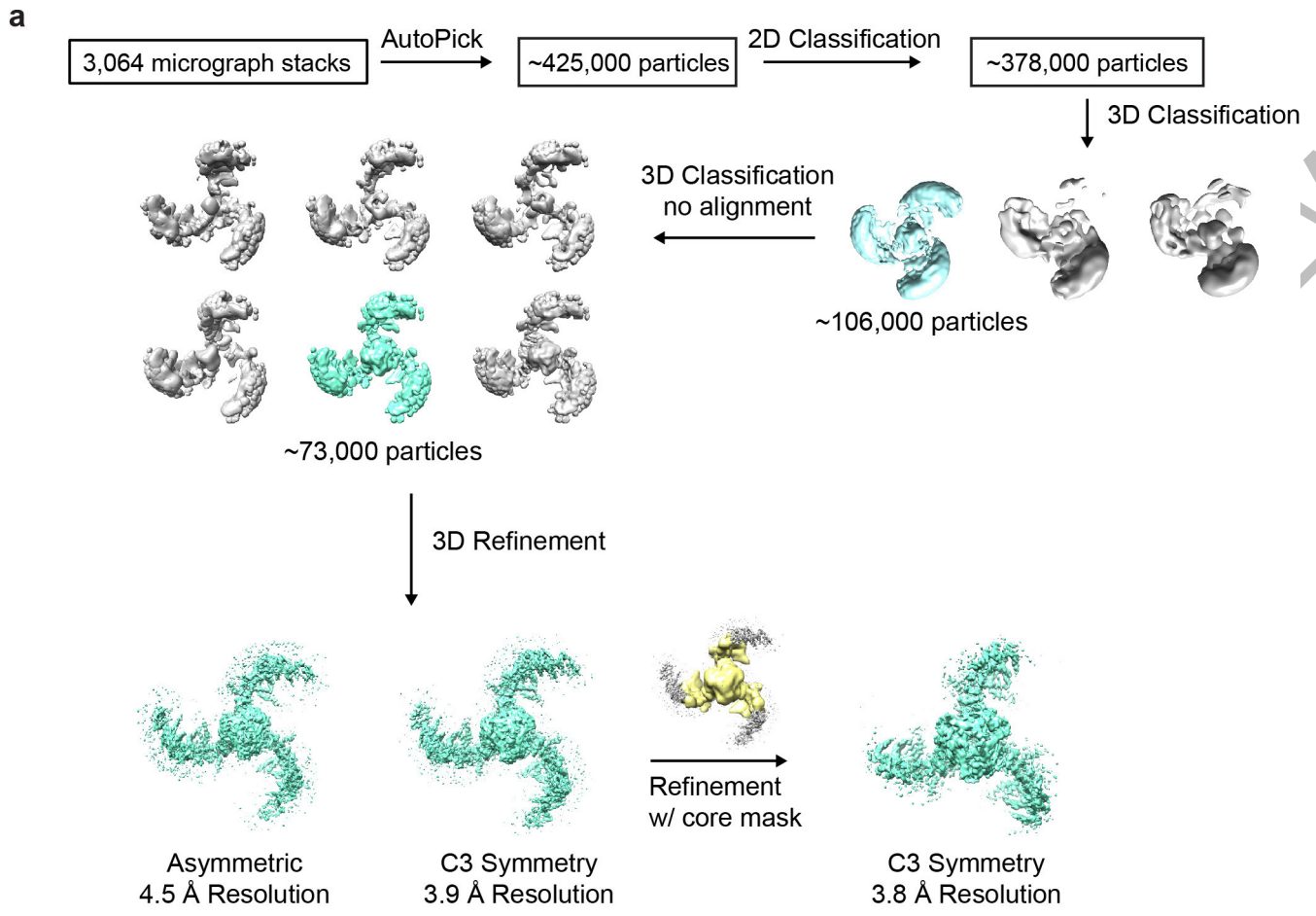
© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

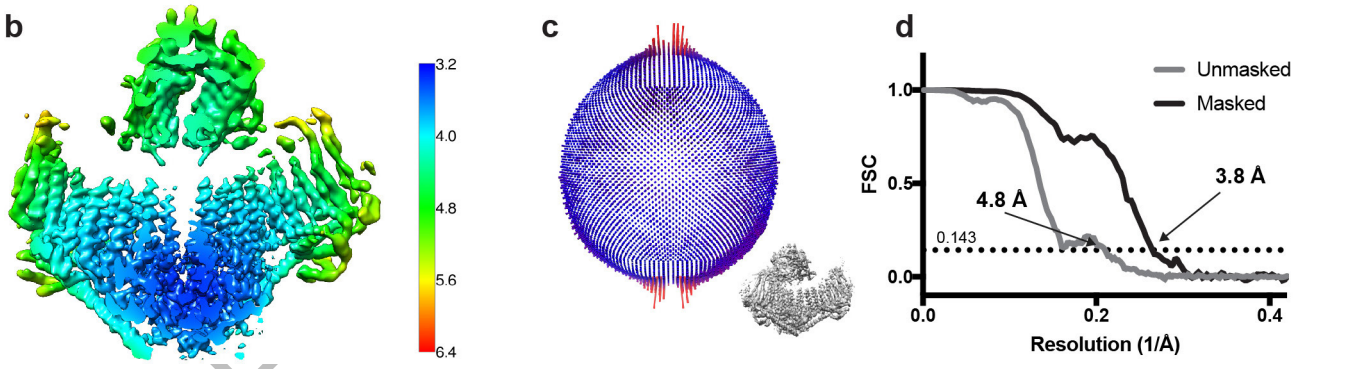


Extended Data Figure 1 | Purification and EM analysis of mPiezo1. **a**, Preparative gel filtration chromatogram of mPiezo1 after affinity purification and proteolytic removal of GST tag. **b**, SDS-PAGE analysis of mPiezo1 following affinity purification (Pre-SEC) and after subsequent SEC step (SEC Fractions). Purifications of mPiezo1 have been repeated more than three times with similar results. **c**, aligned micrograph of purified mPiezo1 embedded in a thin layer of vitrified ice. Scale bar represents 100 nm. **d**, Representative 2D classes of Piezo1 showing different particle orientations.

ACCELERATED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

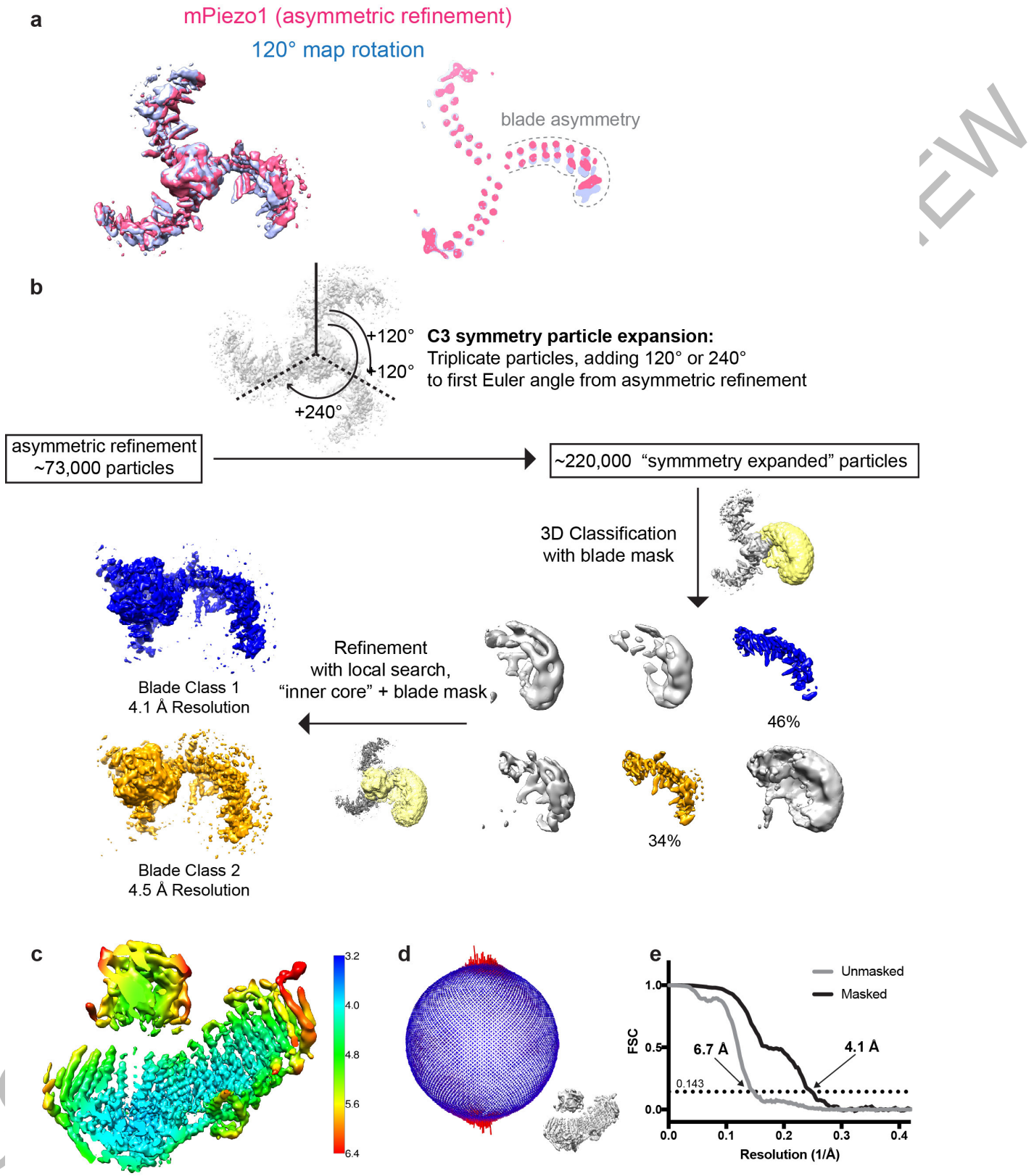




Extended Data Figure 2 | Classification and refinement Piezo1 core. **a,** Data processing flow chart. **b**, local resolution maps calculated by the locres program in RELION 2.0. Color key for local resolution (in Å) is shown. **c**, angular distributions of the particles after the final step of refinement in RELION. The radius of the sphere is proportional to the number of particles with a given orientation. Only one third of the sphere is shown due to applied C3 symmetry. **d**, FSC plots calculated using relion_postprocess for unmasked maps and maps with a soft mask applied to remove the contribution of scattered detergent density.

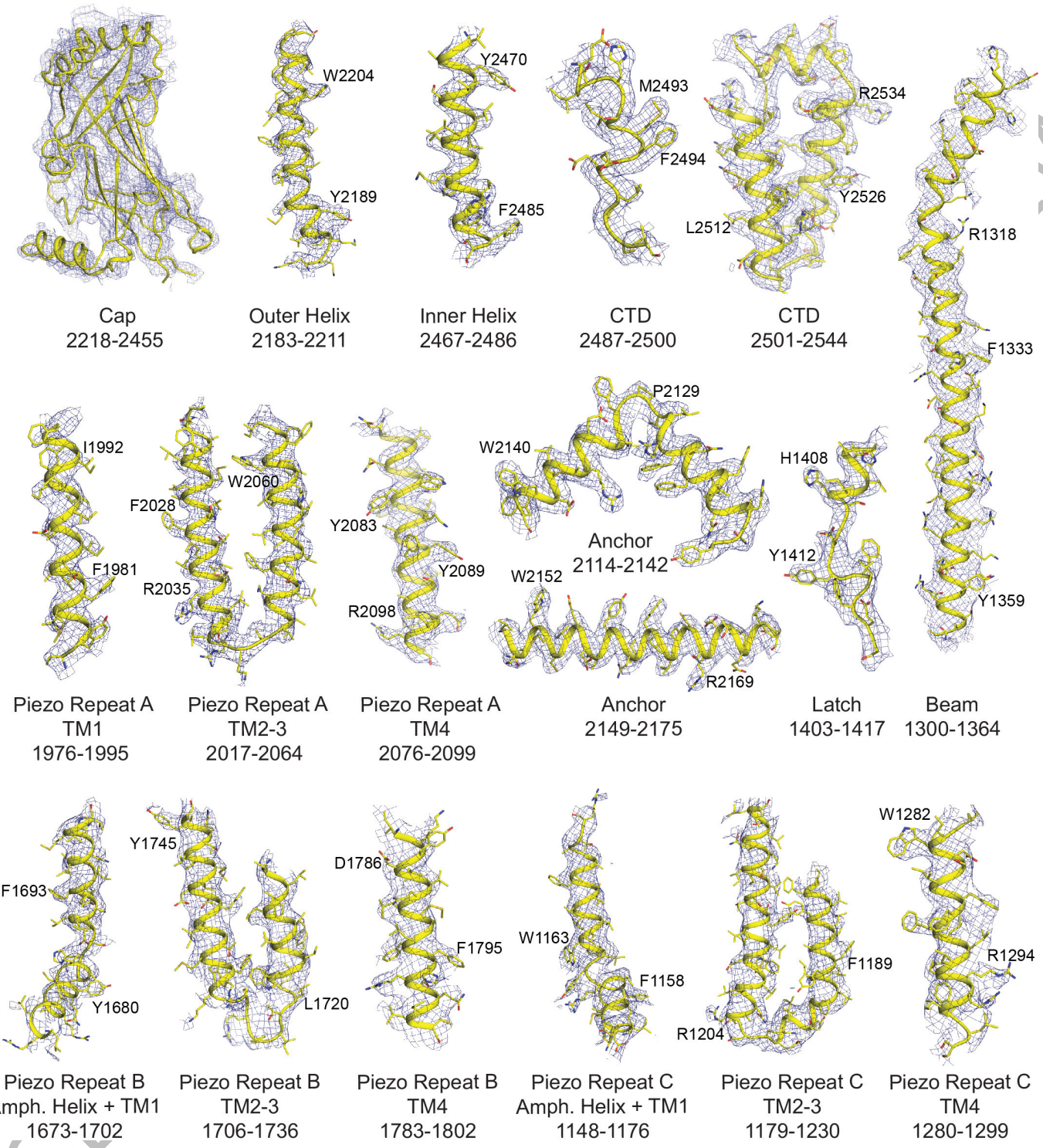
ACCEPTED MANUSCRIPT

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



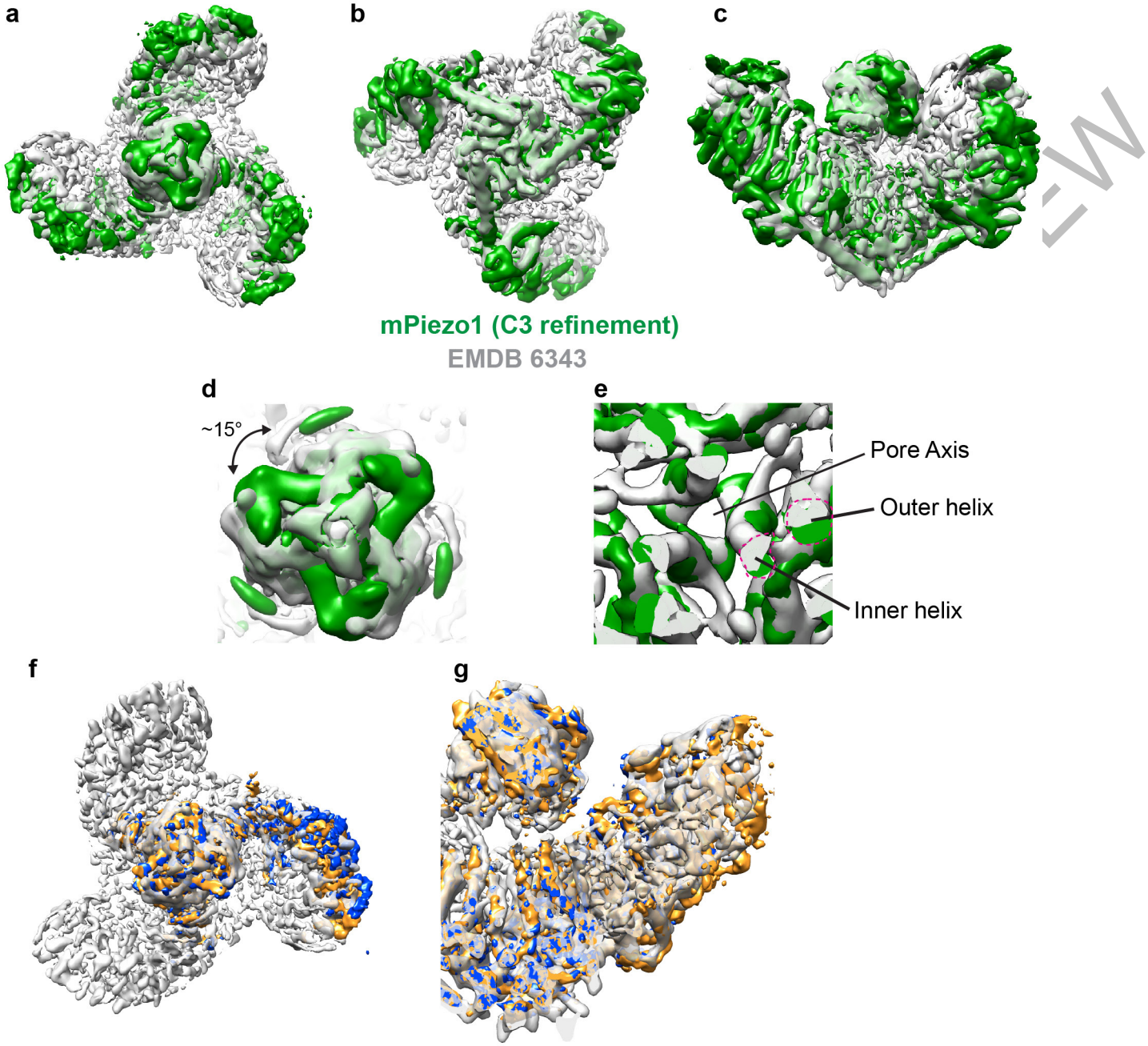
Extended Data Figure 3 | Asymmetry of Piezo1 and blade-focused classification and refinement. a, Duplicate unsharpened maps of Piezo1 refined without symmetry imposed, with the blue map rotated ~120° around the central axis of pseudosymmetry relative to the pink map, then superimposed with the "fit in map" function of UCSF chimera. Left panel shows top view, while right panel shows a horizontal slice through the transmembrane region. **b**, flow chart of propeller blade focused classification and refinement procedure. **c**, local resolution maps calculated by the locres program in RELION 2.0. Color key for local resolution (in Å) is shown and is the same as Extended Data Fig 2b. **d**, angular distributions of the particles after the final step of refinement in RELION. The radius of the sphere is proportional to the number of particles with a given orientation **e**, FSC plots calculated using relion_postprocess for unmasked maps and maps with a soft mask applied to remove the contribution of scattered detergent density.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 4 | Fit of molecular model to electron density. Select regions of the molecular model are shown as yellow cartoon, with superimposed electron density as blue mesh. The density is derived from the C3 symmetry core masked map for the top two rows, and derived from Blade Class 1 for the bottom row.

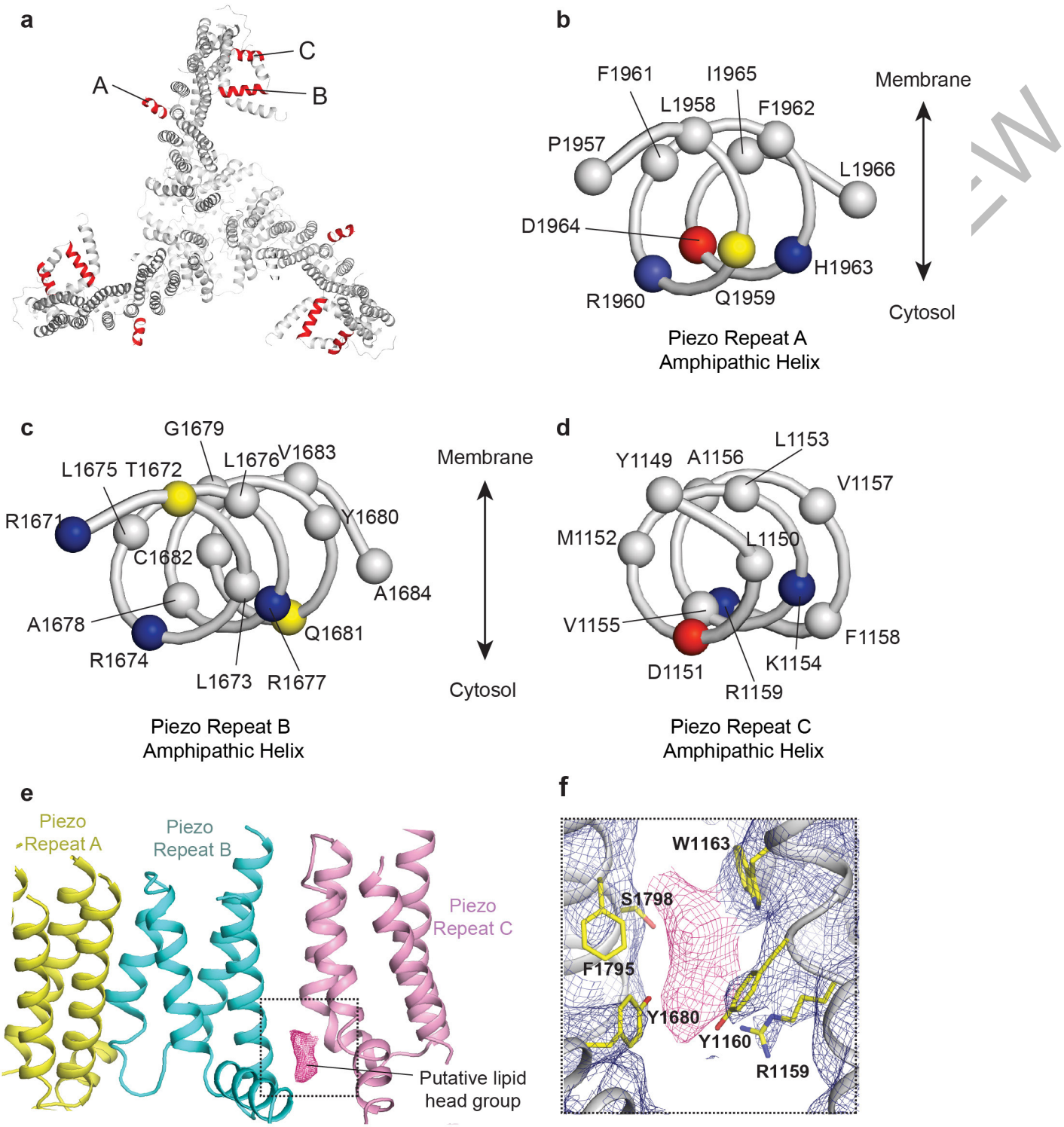
© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 5 | Structural comparisons of current and previous (EMDB 6343) mPiezo1 structures. For each panel, "fit in map" function of UCSF Chimera was used to align uncropped maps. In **a-e**, unsharpened map of C3 refinement is shown. **a-c**, top (**a**), bottom (**b**) and side (**c**) views of current mPiezo1 (green) and EMDB 6343 (gray) structurally aligned. **d** shows expanded top view of the cap domain to illustrate a ~15° rotation of the cap between the two maps. **e** shows expanded sliced top view of pore-proximal transmembrane region. Subtle conformational rearrangements are present between the current and previous structures in the inner and outer helices. **f-g**, top and side views of EMDB 6343 (gray), Blade Class 1 (blue), and Blade Class 2 (orange). Coarsely, the trajectories of the propeller blades are similar across the maps in both the membrane-parallel and membrane-normal directions.

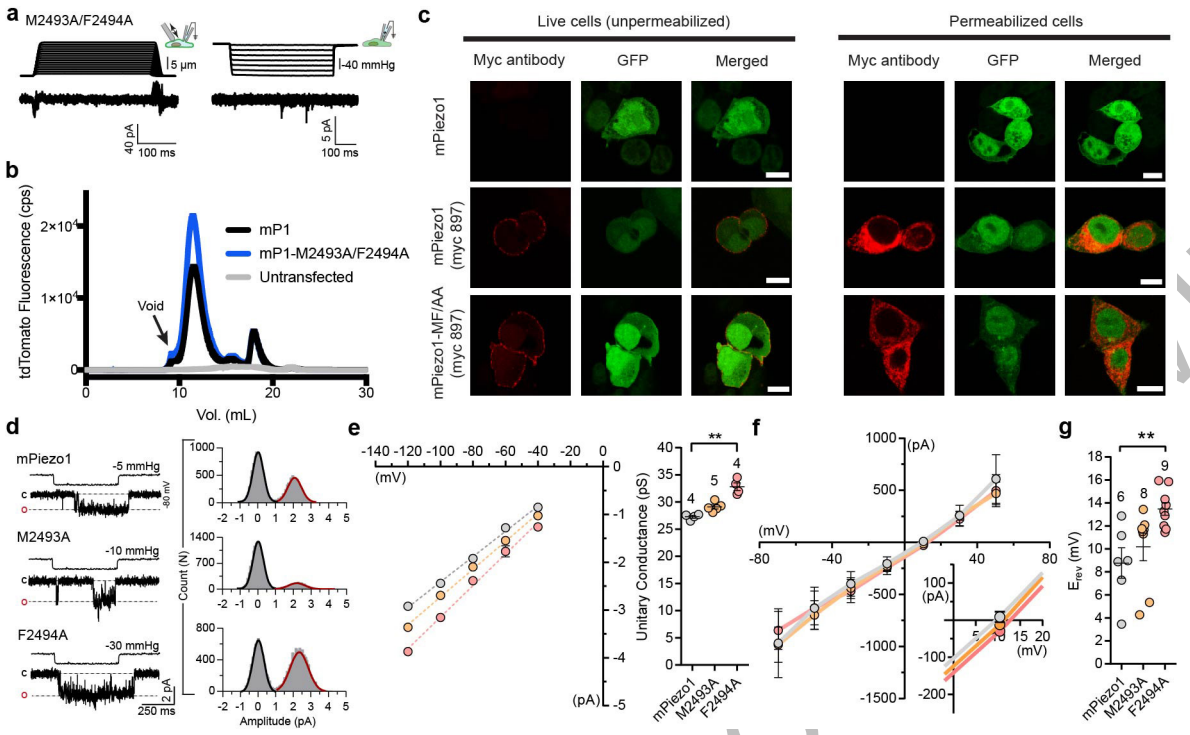
ACCEPTED MANUSCRIPT

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 6 | Structural features of Piezo repeats. a, top view of Piezo1 depicted as cartoon, with amphipathic helices from Piezo Repeats A, B, C shown in red. Extracellular cap domain is omitted for clarity. **b**, **c**, **d**, ribbon models of amphipathic helices from Piezo Repeat A (**b**), Piezo Repeat B (**c**), and Piezo Repeat C (**d**). Cα positions are shown as spheres and hydrophobic amino acids are colored gray, basic amino acids are colored blue, acidic amino acids are colored red, and polar uncharged amino acids are colored yellow. **e,** cartoon model of a portion of the Piezo1 propeller blade, with Piezo Repeats colored separately. A strong density peak putatively representing a lipid head group is shown as pink mesh. **f,** Expanded view of the putative lipid binding site sandwiched between Piezo repeats B and C (helices B4 and C1). Electron density (blue mesh) is superimposed onto the molecular model, and the putative lipid head group is shown as pink mesh. Side chains contributing to the binding pocket are shown as yellow sticks and labeled.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 7 | Expression and functional properties of Piezo1 mutants. a, HEK293T P1KO cells expressing mPiezo1 M2493A/F2494A do not show discernible currents activated by indentation with blunt pipette (left) or stretch (right). **b,** FSEC traces of untransfected HEK293F cells (gray) or cells expressing C-terminal tdTomato fusions of wild-type mPiezo1 or mPiezo1 M2493A/F2494A injected into Superose 6 increase column. Similar to wild-type, the double mutant displays a single dominant peak eluting shortly after void volume, indicating proper trimeric expression. Data from one independent experiment. **c,** Representative images of Myc labeling in mPiezo1 IRES-GFP (top row), mPiezo1-myc IRES-GFP (middle row) or mPiezo1-myc M2493A/F2494A IRES-GFP (bottom row) transfected HEK293T P1KO cells. Myc tags were inserted at position 897. Immunostaining was performed before (left panels) or after (right panels) cell permeabilization. mPiezo1-myc M2493A/F2494A is labeled at the surface in live cells similar to mPiezo-1 myc, indicating that surface expression is preserved in the nonfunctional mutant. Scale bar, 10 µm. Experiments were repeated 3 times with reproducible results. **d,** Left, representative traces of stretch activated single channel currents recorded (-80mV) from HEK293T P1KO cells expressing wild type, M2493A or F2494A mPiezo1 channels (traces displayed after applying 1kHz digital filter). The corresponding pressure stimulus used to elicit the response is illustrated above the current trace. For each condition, the amplitude histogram for the corresponding trace is depicted on the right. The Gaussian fit for the closed (black curve) and open (red curve) components is overlaid on the histograms. **e,** Left, average I-V relationship of stretch activated single channel currents from wild type (N = 4), M2493A (N = 5) or F2494A (N = 4) mPiezo1 channels. Amplitude was measured as a difference in Gaussian fits of full-trace histograms. Right, average unitary conductance calculated from the slope of linear regression line fit to individual cells in each condition. $^{**}P < 0.01$ One-way ANOVA with Dunn's comparison relative to mPiezo1 **f,** Average I-V relationship curves of MA currents recorded from wild type (N = 6), M2493A (N = 8) or F2494A (N = 9) mPiezo1 channels with 150mM CsCl-based intracellular solution and 100mM $CaCl_2$-based extracellular solution. Currents were elicited from -69.6 to 50.4 mV (Δ20 mV). Inset, expanded view of curves between 5 and 15 mV. Values are mean ± s.e.m. **g**, Average reversal potential from individual cells;mPiezo1: 8.8 ± 1.3 mV (n = 6), M2493A: 10.2 ± 1.2 mV (n = 8) and F2494A: 13.5 ± 0.5 mV (n = 9). $^{**}P = 0.0063$ One-way ANOVA with Dunn's comparison relative to mPiezo1. N = X individual cells.

ACCELERATED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | EM data collection, data processing, model refinement and validation

	mPiezo1 Core (EMDB-7128) (PDB 6BPZ)	mPiezo1 Blade Class 1 (EMDB-7128)
Data collection and processing		
Magnification	29,000	29,000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	60	60
Defocus range (μm)	0.8-2.2	0.8-2.2
Pixel size (Å)	1.03	1.03
Symmetry imposed	C3	C1*
Initial particle images (no.)	425,103	425,103
Final particle images (no.)	72,627	102,289*
Map resolution (Å)	3.8	4.1
FSC threshold	0.143	0.143
Refinement		
Map sharpening *B* factor (Å^2)	-118	-152
Model composition		
Protein residues	2700	
Ligands	0	
R.m.s. deviations		
Bond lengths (Å)	0.01	
Bond angles (°)	1.52	
Validation		
MolProbity score	1.68	
Clashscore	4.1	
EMRinger score	1.62 2.54†	
Poor rotamers (%)	0.27	
Ramachandran plot		
Favored (%)	91.8	
Allowed (%)	8.0	
Disallowed (%)	0.2	

*Particles are C3 symmetry expanded (triplicated, adding 120° or 240° to first euler angle from unmasked assymetric refinement, then classified with mask around single blade before refinement).
†EMRinger score calculated for "inner core", which includes Piezo Repeat A, anchor, inner helix, outer helix, CTD, and latch domains.

ACCELERATED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | MA current properties of Piezo1 constructs

Construct	Whole-cell/Poke				Cell-attached/Stretch				
	Imax (pA)	Threshold (μm)	Tau Inactivation (ms)	E_{rev} (mV)	Imax (pA)	P_{50} (mmHg)	Tau Inactivation* (ms)	%SS*	Unitary Conductance (pS)
mPiezo1	2398 ± 728 (7)	3.8 ± 0.31 (7)	13 ± 1 (7)	8.7 ± 1.3 (6)	77 ± 18 (6)	30.7 ± 5.5 (5)	37.6 ± 6 (6)	10.4 ± 1.8 (6)	27.3 ± 0.3 (4)
M2493A	3120 ± 607 (8)	3.6 ± 0.3 (8)	19.6 ± 1.2 (8)	10.2 ± 1.2 (8)	232 ± 73 (7)	47.2 ± 8.7 (7)	68.6 ± 6.3 (7)	11.5 ± 1.3 (7)	29.0 ± 0.4 (5)
F2494A	4172 ± 806 (8)	3.4 ± 0.3 (8)	48.8 ± 6 (8)	13.5 ± 0.5 (9)	122 ± 29 (7)	35.0 ± 1.4 (6)	95.5 ± 15 (7)	30.4 ± 3.6 (7)	32.8 ± 0.7 (4)
M2493A, F2494A	Nonfunctional (11)				Nonfunctional (10)				

*Tau inactivation and percent steady state current (%SS) measured at -60 or -70 mmHg and -60 mmHg, respectively. %SS was measured as the ratio of Imax to residual current at the end of the 500 ms pressure stimulus. Values indicate mean ± s.e.m. Numbers in parenthesis indicate Ns.

ACCELERATED ARTICLE PREVIEW

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ARTICLE

doi:10.1038/nature25743

Structure and mechanogating mechanism of the Piezo1 channel

Qiancheng Zhao[1,2,3]*, Heng Zhou[1,2]*, Shaopeng Chi[1,2,3]*, Yanfeng Wang[1,2,3]*, Jianhua Wang[4], Jie Geng[1,2,3], Kun Wu[1,2,3], Wenhao Liu[1,2,3,5], Tingxin Zhang[1,2,3,5], Meng-Qiu Dong[4], Jiawei Wang[1], Xueming Li[1,2] & Bailong Xiao[1,2,3]

The mechanosensitive Piezo channels function as key eukaryotic mechanotransducers. However, their structures and mechanogating mechanisms remain unknown. Here, we determine the three-bladed, propeller-like cryo-EM structures of mouse Piezo1 and functionally reveal its mechanotransduction components. Despite the lack of sequence repetition, we identify 9 repetitive units constituted of 4 transmembrane (TM) helices each, termed THUs, assembling into the highly curved blade-like structure. The last TM encloses a hydrophobic pore, followed by three intracellular fenestration sites and side portals comprising pore-property-determining residues. The central region forms a 90 Å-long intracellular beam-like structure, which undergoes a lever-like motion for connecting THUs to the pore via the interfaces of the C-terminal domain, anchor-resembling domain and outer helix. Deleting extracellular loops in the distal THUs or mutating single residues in the beam impairs the mechanical activation of Piezo1. Overall, Piezo1 possesses an unprecedented 38-TM topology and designated mechanotransduction components, enabling a lever-like mechanogating mechanism.

The evolutionarily conserved Piezo proteins, including Piezo1 and Piezo2, have been established as the long-sought-after mechanosensitive (MS) cation channels in mammals[1–4]. Piezo channels play critical roles in various mechanotransduction processes[5,6]. For instance, Piezo1 expressed in the blood vessels is critical for sensing blood flow-associated shear stress for proper blood vessel development[7,8], while Piezo2 mediates touch[9–12], proprioception[13], airway stretching and lung inflation[14]. Furthermore, mutations in Piezo genes have been linked to multiple hereditary human diseases involving mechanotransduction[15].

Piezo proteins encode a distinct class of membrane proteins with numerous transmembrane (TM) helices[1,16]. The full-length mouse Piezo1 (mPiezo1) of 2547 residues, heterologously expressed in HEK293T cells and purified[2], has been shown to mediate mechanosensitive cation currents when reconstituted into lipid bilayers[2,17], demonstrating its ability to form intrinsically mechanosensitive cation channels. In cells, Piezo channels can respond to various forms of mechanical stimulation, including poking, stretching and shear stress[1,8,18].

A medium-resolution cryo-electron microscopy (cryo-EM) structure of mPiezo1 has been previously determined, revealing its three-bladed, propeller-shaped trimeric architecture[3]. Combining structural and functional characterizations, we have proposed that the complex mPiezo1 might be divided into the central ion-conducting pore module and the peripheral blade-like mechanotransduction modules[3,4,19]. However, only 14 apparent TMs in each subunit were resolved, and an alanine model of 492 residues together with 227 residues of the C-terminal extracellular domain (CED) resolved by X-ray crystallization was built[3]. Thus, we set up to determine high-resolution structures of mPiezo1 and to functionally reveal its key mechanotransduction components.

Structure determination of mPiezo1

On the basis of analyzing a total of 10171 micrographs with ~2.8 million particles, we determined a structure of mPiezo1 to an overall resolution of 3.97 Å (Extended Data Fig. 1 and 2, Supplementary Data Fig. 1 and Table 1). Overall, the central region including the intracellular beam and the TM region, which contains the 14 TMs resolved in the previously determined structure[3], has clearly better resolution (3 - 4 Å) than the distal blades and the central cap (4 - 7 Å). We identified 12 additional TMs residing in the distal blade (Fig. 1b), which became better resolved when the peripheral blades were subjected to local refinement (Extended Data Fig. 2 and 3a). The connections between the N- and C-termini of the CED and the OH and IH were also better resolved after local refinement (Extended Data Fig. 2 and 3b), allowing assignment of the OH and IH from each subunit. We built and refined a structure model for mPiezo1, which supports important assignments to the major structural domains as shown in Fig. 2, Extended Data Fig. 4 and Supplementary Data Fig. 1. Consistent with the previous structure[3], mPiezo1 possesses a three-bladed, propeller-shaped architecture, comprising the central cap, three peripheral blades and three long intracellular beams (Fig. 1a). The diameter from the top view and the axial height of the structure are 185 Å and 140 Å, respectively (Fig. 1a, middle panel).

The unusually curved peripheral TM blade

24 peripheral TMs in each subunit are organized into 12 parallel pairs (Fig. 2a), extending from the central OH-IH pair to the periphery in a highly curved configuration (Fig. 1b and 2a). When viewed perpendicularly to the plasma membrane (PM) plane, each TM blade twists in a clockwise manner to form a half-circled superhelical structure, resembling the blade of a typical propeller (Fig. 1). The proximal TM25-36 and peripheral TM13-24 appear to segregate into two TM segments, which form a 100° angle (Fig. 1b, left panel). Strikingly, when viewed parallel to the PM plane, the peripheral TM13-24 form a 140° angle relative to the proximal TM25-36 (Fig. 1b, middle panel), which might reside in the normal PM plane. Such an organization appears to place TM25-36 in a normal, while the TM13-24 in a highly curved membrane plane. However, it remains possible that the curved organization of the

[1]School of Pharmaceutical Sciences or Life Sciences. [2]Tsinghua-Peking Joint Center for Life Sciences. [3]IDG/McGovern Institute for Brain Research, Tsinghua University, Beijing, 100084, China. [4]National Institute of Biological Sciences, Beijing, 102206, China. [5]Joint Graduate Program of Peking-Tsinghua-NIBS, School of Life Sciences, Tsinghua University, Beijing 100084, China.
*These authors contributed equally to this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ACCELERATED ARTICLE PREVIEW

TMs might be due to removal of the purified Piezo1 proteins from the membrane. Characteristically, TM13, 17, 21, 25 and 29 form L-shaped helical structures with TM13$^{pre\alpha1}$, TM17$^{pre\alpha1}$, TM21$^{pre\alpha2}$, TM25$^{pre\alpha2}$ and TM29$^{pre\alpha2}$, respectively, which are positioned parallel to the membrane and collectively form an intracellular helical layer immediately underneath the membrane (Fig. 2a and Extended Data Fig. 3a), which might help to stabilize the curved TM blade in the membrane. The highly curved TM blade and the intracellular helical layer might represent unique structural features not only for mechanosensing and transduction but also for inducing local membrane curvature.

The THU repeats and a 38-TM topology model

The assignment of residues into the peripheral TM25-36 revealed 3 repetitive folds containing 4-TMs each, which we term transmembrane helical units (THU) (Fig. 2 and Extended Data Fig. 4, 5). Within each THU, the first two TMs, closely spaced with a short extracellular linker, are folded toward the central end (referring to the OH-IH pair) of the TM blade. The second TM folds backward to the distal end and connects to the third TM, which connects to the fourth TM through an extracellular linker of 30 - 50 residues. Two consecutive THUs are connected through a relatively long intracellular loop, which spans all 8 TMs and consequently arranges the 4 pairs of TMs in series (Fig. 2). TM25-28, TM29-32 and TM33-36 form THU7, THU8 and THU9, respectively (Fig. 2). The loop that links TM27 and TM28 was experimentally tested to be located extracellularly (Extended Data Fig. 6).

Based on the folding characteristics of the THUs, topology and secondary structure prediction, and cross-linking results (Extended Data Fig. 4, 5, 7 and Supplementary Data Fig. 1), the structurally revealed TM13-16, 17-20 and 21-24 were assigned to THU4, 5 and 6, respectively. These 6 tandem THUs constitute the highly curved TM blade (Fig. 2). Many topology prediction programs have predicted 12 TMs in the unresolved N-terminal region of ~500 residues (Extended Data Fig. 5). Each 4-TM of the predicted 12 TMs follows the typical folding features of the THU. In line with the prediction, the second extracellular loops in each THU have been experimentally verified to be extracellularly located[20] (Fig. 2b). Furthermore, we identified residues from the linkers of TM4-5 (THU1-2), TM8-9 (THU2-3), TM16-17 (THU4-5) that cross-linked with residues in the intracellular loop of TM28-29 (THU7-8) or TM32-33 (THU8-9) (Extended Data Fig. 7 and Fig. 2b), indicating the intracellular localization of these loops. Based on these lines of evidence, we assign the N-terminal 12 TMs to THU1-3. Therefore, the peripheral 36 TMs are folded into 9 tandem repetitive THUs to form the unique TM blade. Together with the OH and IH, Piezo1 might possess an unprecedented 38-TM topology with a total of 114 TMs in the trimeric channel complex.

The intracellular beam

Three characteristic long density rods ~90 Å in length are exposed on the intracellular surface of THU7-9 and CTD and are termed the beams (Fig. 1, 2). We were able to assign H1300- S1362 to the beam, which is apparently kinked at residues A1316/S1317 (Extended Data Fig. 4). This assignment is in line with the predicted continuous long helix immediately following TM28 (Supplementary Data Fig. 1) and supported by the CXMS results (Extended Data Fig. 7). The THU7-8 loop is the largest intracellular loop of Piezo1, containing ~390 residues. According to the structure, this loop starts at the distal end of the beam, extends ~90 Å into the center of the complex to interact with the CTD, and then folds back to the distal end of the beam prior to connecting to TM29 (Fig. 2). The organization of this loop might render the beam an ideal structure for mechanical transmission from the distal THUs to the central ion-conducting pore.

The beam-CTD-anchor-IH/OH relaying interfaces

CTD$^{\alpha1-3}$ forms a triangular plane parallel to the plasma membrane, positioned immediately above the proximal end of the beam (Fig. 2a and 3a). The beam-facing side of the triangular CTD plane is clearly separated into two surfaces with negative and positive electrostatic potential (Fig. 3d). The beam crosses through the boundary of the two surfaces, forming interactions with both CTD$^{\alpha1}$ and CTD$^{\alpha2}$. The disease-associated residue R1353 might form a salt bridge with E2518 (Fig. 3d, e).

The anchor domain, uniquely positioned between the OH-IH pair and the CTD plane, results in a swap of the OH-CED-IH into the neighboring subunit (Fig. 1 and 3a). Anchor$^{\alpha1-2}$ forms an inverted v-shaped structure that penetrates into the inner leaflet of the membrane, containing the evolutionarily conserved motif among Piezo homologs, P2129F2130(X2)E2133(X6)W2140[21]. P2129, F2130, and E2133 are in close proximity to the intracellular end of the pore-lining IH (Fig. 3a, b), which may stabilize the integrity of the ion-conducting pore. Indeed, although E2133 is not located in the ion-conducting pathway, its mutations affect the pore properties[20]. Anchor$^{\alpha2}$ lies parallel to the inner portion of the OH, which is kinked at residue I2203, forming hydrophobic interactions (Fig. 3a, b). The upper portion of the OH above the kink lies parallel to the IH and also forms extensive hydrophobic interactions (Fig. 3a, b). Thus, the OH might help to stabilize the IH-encompassing pore.

The long anchor$^{\alpha3}$, in parallel with the membrane plane, forms a hydrophobic interface with the CTD (Fig. 3a, c). It connects to the OH through the lysine-rich anchor-OH-linker (K2174YPQPKGQKKKK2184), which forms a tripartite interaction with the polar residue-rich anchor$^{\alpha2-3}$-turn (T2143DTTTLS2148), and the glutamate-rich region of the CTD (R2514ETRELALEEE2524) (Fig. 3a, e). Several residues, including T2143, KKKK(2182-K2185), R2514, E2522, and E2523, are disease-associated residues. Furthermore, we have found that the anchor-OH-linker is important not only for mechanogating but also for mediating SERCA2 interaction and the resulting inhibition of Piezo1[22].

The ion-conducting pathway

The OH-CED-IH-CTD structure trimerizes to form the well-resolved central pore module[3,4]. Subjecting the structure of the pore module to the online PPM server[23] revealed that the entire IH spanning from Y2464 to F2485 defines the transmembrane region of ~29 Å thick (Fig. 4a, b). The central solvent-accessible pathway can be separated into the extracellular vestibule (EV) above the membrane, the membrane vestibule (MV) within the membrane and the intracellular vestibule (IV) immediately below the membrane (Fig. 4a, c). The IH lacks negatively charged residues and encloses a hydrophobic transmembrane pore with pore-facing residues shown in Fig. 4a, b, e. V2476 forms a constriction point in the middle of the MV (Fig. 4a-c). Notably, the transmembrane pore is not completely sealed from the membrane by the three IHs (Fig. 4d, e). Two neighboring IHs and one OH form a hydrophobic groove open to the transmembrane pore (Fig. 4d, e). This feature raises an intriguing possibility that membrane lipids may affect the ion permeation and gating of Piezo1.

Intriguingly, while the top of the EV is sealed and the bottom end of the IV is constricted (Fig. 4a, c), both EV and IV have large fenestration sites immediately above and below the membrane, respectively (Fig. 4d). The patch of negatively charged residues DEEED(2393-2397), located right above the extracellular fenestration sites (Fig. 4d), plays a critical role in controlling efficient ion conduction and the selection of cations over anions[3,4]. These features suggest that cations might enter the ion-conducting pathway through the extracellular fenestration sites. The intracellular fenestration sites have a size of ~10 x 15 Å, through which the IV connects to three ~8 Å-wide side portals with negative electrostatic potentials, comprising E2487 from one subunit and E2495 and E2496 from a neighboring subunit (Fig. 4d, f). Mutating these residues affects the unitary conductance, Ca^{2+} permeability and ruthenium red responsiveness of mPiezo1[4]. Given that E2495 and E2496 are not located along the central pore axis but rather on the side portals (Fig. 4d, f), we propose that the three intracellular fenestration sites and the connecting side portals might represent the intracellular cation permeation pathways. Nevertheless, it remains to be determined

ACCEPTED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

whether the top seal of the EV and the bottom constriction site of the IV might be gated for allowing ion conduction. Previous studies have indicated that the diameter of the ion-conducting pathway upon opening is larger than 8 Å[24]. Thus, the observed constriction sites in the transmembrane pore indicate that the structure represents a closed conformation (Fig. 4a, c).

Motion features of Piezo1

Comparing the 9 distinct structures derived from symmetry-free classification revealed conformational changes at the peripheral blades, the beam and the cap (Extended Data Fig. 8). Intriguingly, as shown in Fig. 5a, each subunit within the trimeric complex might be able to move independently. The independent motion of the individual subunit might be functionally relevant, as a given force in the membrane is unlikely to be evenly distributed among the three subunits unless the force directly acts on the central cap region. We further compared two structures of 4.51 Å and 4.40 Å resolution, resulting from class 3 and class 6 of the 9 classified structures with C3-symmetry, respectively (Extended Data Fig. 2). Superimposition of the two maps shows that the IHs, anchor and CTD were nearly unchanged (Fig. 5c, d). By contrast, the cap of class 6 rotates in a clockwise manner relative to that of class 3 (Fig. 5b). The TM blades of class 6 undergo an anticlockwise twist, with a graded decrease in motion from the distal THU4 to the proximal THU9 (Fig. 5b). The intracellular helical layer exhibits a vertical motion (Fig. 5d), which might flatten the curved distal blades and the residing membrane. Notably, the beam displays uneven movement with large motion at the distal beam while subtle movement at the proximal end (Fig. 5d - f). As a whole, the motion feature of the peripheral TMs and the beams is reminiscent of a lever apparatus (Fig. 5f). We next carried out biochemical and functional studies to test the importance of the distal blade and the beam in the mechanical activation of Piezo1.

The extracellular loops in distal THUs are crucial for Piezo1 expression and mechanical activation

We generated a series of deletion mutants in which the extracellular loops of distal THU1-6 were deleted (replaced with a short GGGG linker) one at a time (Fig. 2b). Among the deletion mutants, only ΔL15-16 and ΔL19-20 had comparable expression to that of wild-type mPiezo1 (Extended Data Fig. 9d and Supplementary Data Fig. 2b), and were properly targeted to the PM (Extended Data Fig. 9a-c and Supplementary Data Fig. 2a). Together, these data suggest that the loop regions of TM3-4, TM7-8, TM11-12, and TM23-24 are critical for the proper expression and PM targeting of Piezo1. ΔL15-16 and ΔL19-20, but not other mutants, were functional (Fig. 6a, b, e). However, the poking-induced maximal whole-cell currents (Imax) of ΔL15-16 and ΔL19-20 were only 4.3 ± 1.3% and 3.3 ± 0.8% of the mPiezo1-mediated current, respectively (Fig. 6a, b). Similar results were observed in ΔL15-16- or ΔL19-20-transfected Piezo1 knockout HEK293 cells in which the endogenous Piezo1 gene was disrupted[25] (Extended Data Fig. 9e). Furthermore, the stretching-induced Piezo1-like currents were abolished in cells expressing the two mutants (Fig. 6d, e). Yoda1, a chemical agonist of Piezo1, can evoke Piezo1-dependent Ca^{2+} response and potentiate Piezo1-mediated mechanical currents[26]. Compared to mPiezo1, the two mutants showed comparable Yoda1-induced fold changes of the poking evoked currents (Fig. 6c) and Ca^{2+} responses (Fig. 6f and Extended Data Fig. 9i), suggesting that the two mutants retain normal Yoda1 responsiveness. In line with their drastically reduced mechanically evoked currents in the absence of Yoda1 (Fig. 6a, b), the poking-induced Imax of the two mutants was also smaller than that of mPiezo1 when Yoda1 was present (Fig. 6a and Extended Data Fig. 9g). The inactivation kinetics of the two mutants was similar to that of mPiezo1 (Extended Data Fig. 9f, g), suggesting that deleting the two loops does not affect the inactivation process. Together, these data demonstrate that the two extracellular loops play a critical role in the mechanical activation of Piezo1.

Residues L1342 and L1345 in the beam are required for mechanical activation

Deleting the residues from 1280 to 1360 that contain the beam structure abolished the expression of the resulting mutant protein (Extended Data Fig. 9d), in line with the structural importance of the beam. Interestingly, the beam domain contains a predicted coiled-coil motif, LAQLKRQM (1342-1349), which is located in the proximal end of the beam near the CTD and appears to be stable (Fig. 5f). The L1342A/L1345A double mutant was properly expressed in the PM (Extended Data Fig. 10a, b). However, the mutant exhibited reduced poking-induced currents (Fig. 6g, h). Its Imax reached only ~21% of the mPiezo1-mediated current (Fig. 6h). L1342A/L1345A also exhibited reduced stretching-induced current (~47% of the mPiezo1 current) (Fig. 6i), rightward shifted pressure-current relationship (Fig. 6j), and enhanced pressure required to activate half of the channel (P_{50}) (mPiezo1 vs L1342A/L1345A: -27.1 ± 3.1 vs -46.2 ± 6.4 mmHg) (Fig. 6k). The single-channel conductance (Extended Data Fig. 10d, e) of the mutant was not affected. Furthermore, neither mutating the residue Q1344 (Extended Data Fig. 10f), nor mutating residues in close proximity to L1342 and L1345, including T2103 and R2104 in the TM36-anchor-linker and L2512 and T2516 in the CTD (Extended Data Fig. 10g, h), affected the poking-induced currents. Collectively, these data suggest that L1342 and L1345 are specifically required for the mechanogating of Piezo1.

The ΔL15-16 and ΔL19-20 mutants completely lost their ability to generate stretch-induced currents (Fig. 6d, e), but could produce residual poking-currents (Fig. 6a, b). The L1342A/L1345A mutant has more drastically reduced poking-induced currents than stretching-induced currents (Fig. 6h, i). These observations suggest that different forms of mechanical stimulation may employ discrete molecular bases of Piezo1 for mechanotransduction.

Discussion

Piezo channels serve as a principal type of mechanotransduction channels. Here, via determining the mPiezo1 structures, we have revealed its unique topological features, the ion-conducting pathway, and the lever-like motion of the blade-beam. Functionally, we have identified regions and single residues that are critical for the mechanical activation of Piezo1. We propose that Piezo1 might employ its characteristically curved blades and the long beams with the L1342/L2345 as a pivot to form a lever-like apparatus. Such a lever-like mechanotransduction mechanism might enable Piezo channels to effectively convert a large conformational change of the distal blades to a relatively slight opening of the central pore, allowing cation-selective permeation. Three sets of such lever-like apparatus are further assembled into a gigantic propeller-like machinery, which might confer a coordinated mechanosensitivity. Thus, different classes of mechanosensitive ion channels, such as the bacterial mechanosensitive channels with large conductance (MscL), the mechanosensitive K2P channels and the NOMPC mechanotransduction channels, employ distinct structural and functional mechanisms for mechanotransduction[27–29].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 October 2017; accepted 12 January 2018.
Published online 22 January 2018.

1. Coste, B. *et al.* Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science* **330**, 55–60, doi:10.1126/science.1193270 (2010).
2. Coste, B. *et al.* Piezo proteins are pore-forming subunits of mechanically activated channels. *Nature* **483**, 176–181, doi:10.1038/nature10812 (2012).
3. Ge, J. *et al.* Architecture of the mammalian mechanosensitive Piezo1 channel. *Nature* **527**, 64–69, doi:10.1038/nature15247 (2015).
4. Zhao, Q. *et al.* Ion Permeation and Mechanotransduction Mechanisms of Mechanosensitive Piezo Channels. *Neuron*, doi:10.1016/j.neuron.2016.01.046 (2016).

ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

5. Geng, J., Zhao, Q., Zhang, T. & Xiao, B. In Touch With the Mechanosensitive Piezo Channels: Structure, Ion Permeation, and Mechanotransduction. *Curr Top Membr* **79**, 159–195, doi:10.1016/bs.ctm.2016.11.006 (2017).
6. Ranade, S. S., Syeda, R. & Patapoutian, A. Mechanically Activated Ion Channels. *Neuron* **87**, 1162–1179, doi:10.1016/j.neuron.2015.08.032 (2015).
7. Ranade, S. S. *et al.* Piezo1, a mechanically activated ion channel, is required for vascular development in mice. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 10347–10352, doi:10.1073/pnas.1409233111 (2014).
8. Li, J. *et al.* Piezo1 integration of vascular architecture with physiological force. *Nature*, doi:10.1038/nature13701 (2014).
9. Maksimovic, S. *et al.* Epidermal Merkel cells are mechanosensory cells that tune mammalian touch receptors. *Nature* **509**, 617–621, doi:10.1038/nature13250 (2014).
10. Woo, S. H. *et al.* Piezo2 is required for Merkel-cell mechanotransduction. *Nature*, doi:10.1038/nature13251 (2014).
11. Ranade, S. S. *et al.* Piezo2 is the major transducer of mechanical forces for touch sensation in mice. *Nature* **516**, 121–125, doi:10.1038/nature13980 (2014).
12. Ikeda, R. *et al.* Merkel cells transduce and encode tactile stimuli to drive Abeta-afferent impulses. *Cell* **157**, 664–675, doi:10.1016/j.cell.2014.02.026 (2014).
13. Woo, S. H. *et al.* Piezo2 is the principal mechanotransduction channel for proprioception. *Nature neuroscience* **18**, 1756–1762, doi:10.1038/nn.4162 (2015).
14. Nonomura, K. *et al.* Piezo2 senses airway stretch and mediates lung inflation-induced apnoea. *Nature*, doi:10.1038/nature20793 (2016).
15. Alper, S. L. Genetic Diseases of PIEZO1 and PIEZO2 Dysfunction. *Curr Top Membr* **79**, 97–134, doi:10.1016/bs.ctm.2017.01.001 (2017).
16. Kamajaya, A., Kaiser, J. T., Lee, J., Reid, M. & Rees, D. C. The structure of a conserved piezo channel domain reveals a topologically distinct beta sandwich fold. *Structure* **22**, 1520–1527, doi:10.1016/j.str.2014.08.009 (2014).
17. Syeda, R. *et al.* Piezo1 Channels Are Inherently Mechanosensitive. *Cell reports* **17**, 1739–1746, doi:10.1016/j.celrep.2016.10.033 (2016).
18. Cox, C. D. *et al.* Removal of the mechanoprotective influence of the cytoskeleton reveals PIEZO1 is gated by bilayer tension. *Nature communications* **7**, 10366, doi:10.1038/ncomms10366 (2016).
19. Wang, Y. & Xiao, B. The mechanosensitive Piezo1 channel: structural features and molecular bases underlying its ion permeation and mechanotransduction. *The Journal of physiology*, doi:10.1113/JP274404 (2017).
20. Coste, B. *et al.* Piezo1 ion channel pore properties are dictated by C-terminal region. *Nature communications* **6**, 7223, doi:10.1038/ncomms8223 (2015).
21. Prole, D. L. & Taylor, C. W. Identification and analysis of putative homologues of mechanosensitive channels in pathogenic protozoa. *PloS one* **8**, e66068, doi:10.1371/journal.pone.0066068 (2013).
22. Zhang, T., Chi, S., Jiang, F., Zhao, Q. & Xiao, B. A protein interaction mechanism for suppressing the mechanosensitive Piezo channels. *Nature communications* **8**, 1797, doi:10.1038/s41467-017-01712-z (2017).
23. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic acids research* **40**, D370–376, doi:10.1093/nar/gkr703 (2012).
24. Gnanasambandam, R., Bae, C., Gottlieb, P. A. & Sachs, F. Ionic Selectivity and Permeation Properties of Human PIEZO1 Channels. *PloS one* **10**, e0125503, doi:10.1371/journal.pone.0125503 (2015).
25. Lukacs, V. *et al.* Impaired PIEZO1 function in patients with a novel autosomal recessive congenital lymphatic dysplasia. *Nature communications* **6**, 8329, doi:10.1038/ncomms9329 (2015).
26. Syeda, R. *et al.* Chemical activation of the mechanotransduction channel Piezo1. *eLife* **4**, doi:10.7554/eLife.07369 (2015).
27. Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T. & Rees, D. C. Structure of the MscL homolog from Mycobacterium tuberculosis: a gated mechanosensitive ion channel. *Science* **282**, 2220–2226 (1998).
28. Brohawn, S. G., Campbell, E. B. & MacKinnon, R. Physical mechanism for gating and mechanosensitivity of the human TRAAK K+ channel. *Nature* **516**, 126–130, doi:10.1038/nature14013 (2014).
29. Jin, P. *et al.* Electron cryo-microscopy structure of the mechanotransduction channel NOMPC. *Nature* **547**, 118–122, doi:10.1038/nature22981 (2017).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Dr. Nieng Yan for critical discussion, and the Beijing Advanced Innovation Center for Structural Biology for facility and financial support. This work was supported by grant numbers 31630090, 2016YFA0500402, 31422027, 31371118, and 2015CB910102 to B.X.; 31570730, 2016YFA0501102 and 2016YFA0501902 to X.L.; and 21375010 to M.-Q.D., from either the National Natural Science Foundation of China or the National Key R&D Program of China.

Author Contributions Q.Z. performed cloning, protein purification, immunostaining, EM data collection; H.Z. performed EM sample preparation, data collection and analysis; S.C., Y.W., J.G., K.W., W.L., T.Z. carried out biochemical and functional studies; J.W. performed mass spectrometry under the supervision of M.-Q.D.; J.W. built the model; X.L. directed the EM data collection and analysis; B.X. conceived and directed the study, analyzed the structure and assisted in model building, and wrote the manuscript with help from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to B.X. (xbailong@mail.tsinghua.edu.cn) and X.L. (lixueming@mail.tsinghua.edu.cn).

Reviewer Information *Nature* thanks E. McCleskey and the other anonymous reviewer(s) for their contribution to the peer review of this work.

ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

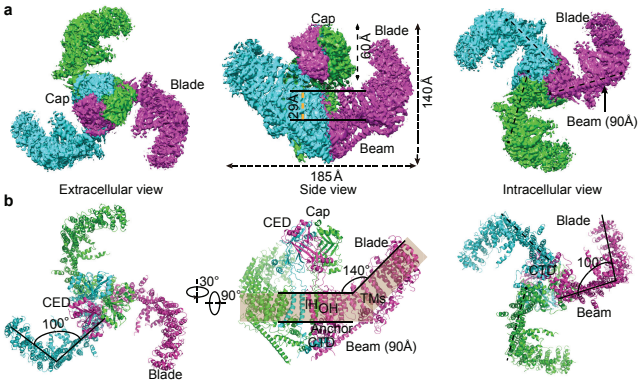


Figure 1 | Overall structure of mPiezo1. a, The indicated view of the sharpened map (6σ contour level) filtered to a resolution of 3.97 Å, with each subunit color-coded and the major domains labeled. **b**, Cartoon models with each subunit color-coded. In the middle panel, the front subunit is omitted for a better view of the curvature of the TMs. The shadow area indicates a planar and a potentially curved membrane plane where TMs reside.

ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

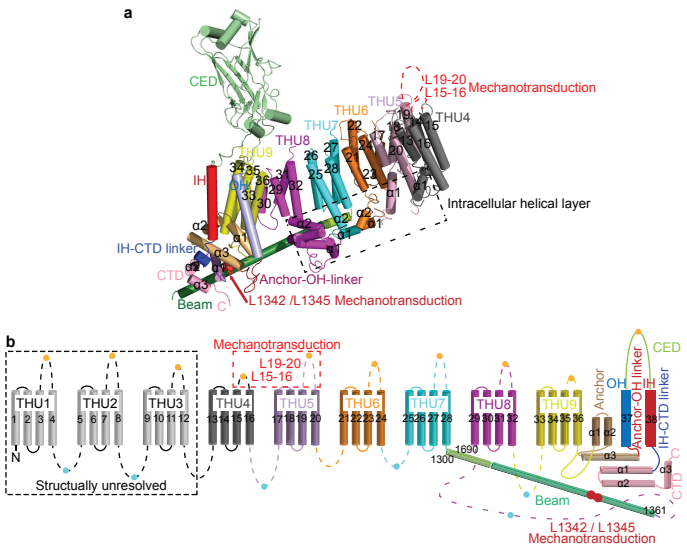


Figure 2 | Repetitive THUs and a 38-TM topology model. a, A cartoon model showing cylindrical helices displays one subunit with individual THUs and major structural domains labeled. The functionally characterized extracellular loops of TM15-16 (L15-16) and TM19-20 (L19-20) are shown in red dashed lines, while residues L1342 and L1345 in the beam are shown by red spheres. **b**, A 38-TM topology model, color-coded to match the cartoon models in **a**. Some intracellular helices shown in **a** are omitted for better illustration of the THUs and beam features. The yellow dots mark the experimentally verified extracellular loops, while the blue dots indicate loops containing residues that can be cross-linked to residues in the intracellular loop of TM32-33.

ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

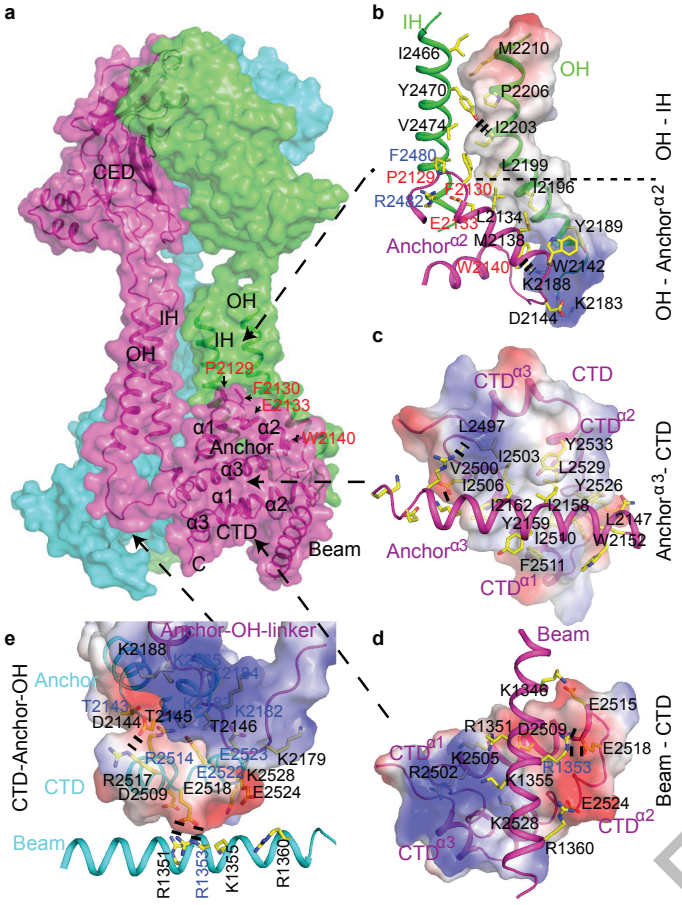


Figure 3 | Beam-CTD-Anchor-OH/IH interfaces. a, Surface representation (transparent) of the segmented map of the indicated structural domains. The highly conserved residues among Piezo homologous are labeled and shown in red. **b** - **e**, The indicated interfaces with one part illustrated in ribbon representation and the other part in surface representation. Dashed lines indicate polar interactions. Blue residues indicate disease-causing residues.

ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

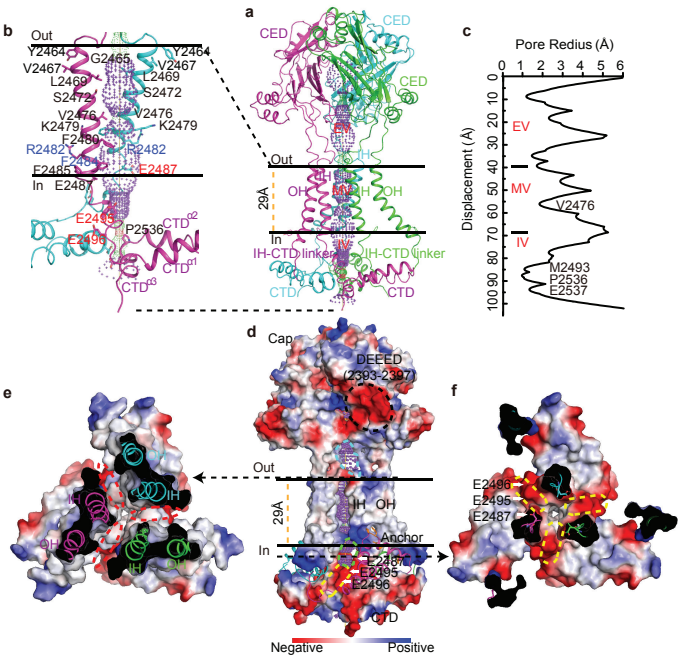


Figure 4 | The ion-conducting pathway. a, Ribbon diagram of the ion conduction pore module. The central solvent-accessible pathway is marked with a dotted mesh generated by the program HOLE (pore radius: red < 1 Å < green < 2 Å < purple). **b**, Diagram showing an enlarged illustration of the IH-encoded and CTD-encoded portion of the ion conduction pathway. The residues in the IH, disease-causing residues (in blue), and residues critical for pore properties (in red) are labeled. **c**, Pore radius along the central axis of the ion conduction pathway. The constriction size at the EV-MV junction is based on an alanine model of the CED-IH linker. **d**, Pore module showing the surface electrostatic potential. The central pathway is marked with a dotted mesh, showing the opening to the membrane. The extracellular and intracellular fenestration sites open to the EV and the IV are marked by cyan and green dashed lines, respectively. The intracellular side portals are marked with the yellow dashed lines. **e**, Diagram showing a cross-section at the arrow indicated plane, showing the hydrophobic transmembrane pore and the side openings to the membrane marked by the dashed red ovals. **f,** Diagram showing a cross-section near the E2487 residue, highlighting the three side portals formed by E2487, E2495 and E2496.

ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

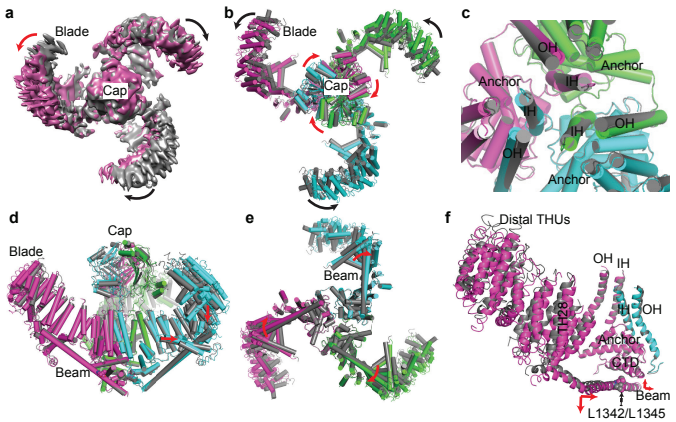


Figure 5 | Motion features. a, Overlay of the class 4 (in gray) and 6 (in purple) structures without symmetry, showing the opposite twisting motion of one blade (marked by the red arrow) compared to the other two blades (marked by black arrows). **b**, **d**, **e** The indicated view of the superimposed class 3 (in gray) and 6 (color-coded) structures with C3 symmetry. **c**, Diagram showing the enlarged top view of the pore by omitting the cap from **b**. **f**, Diagram showing the uneven motion of the beam. The L1342 and L1345 residues are shown in sphere representation. The distal THUs and the central CTD-Anchor-OH/IH relaying interfaces are included to illustrate the lever-like mechanotransduction apparatus constituted by the beam.

ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

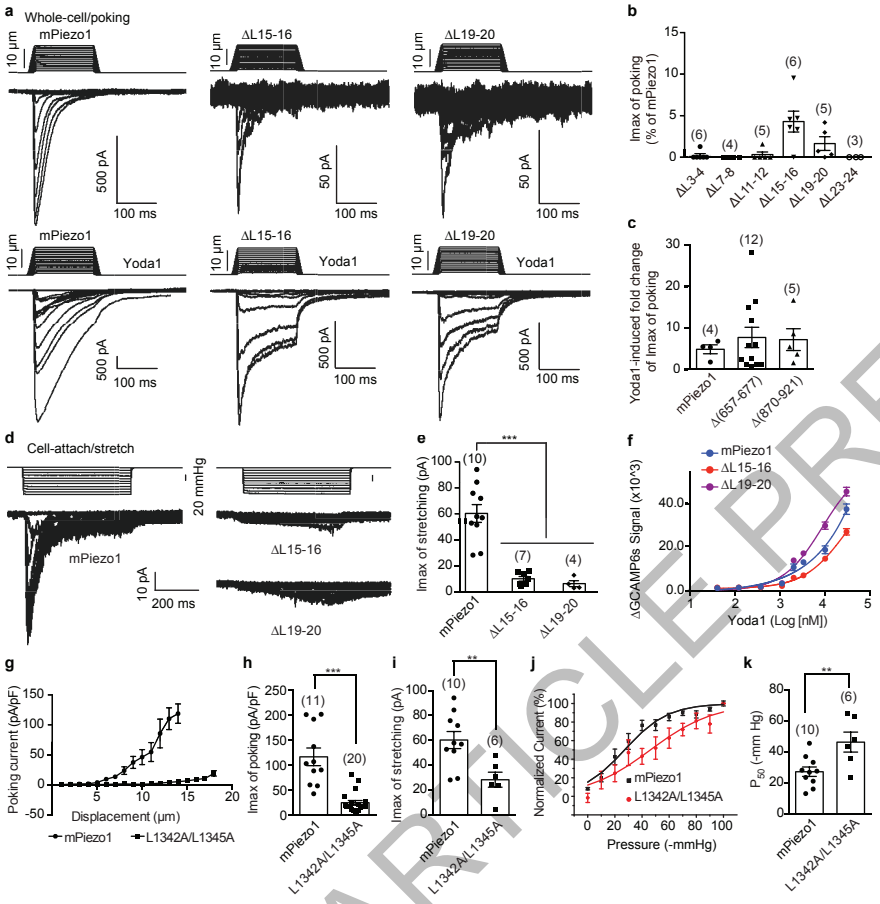


Figure 6 | Regions and residues critical for the mechanical activation of Piezo1. a, Representative traces of poking-induced inward currents at -60 mV in the whole-cell configuration in the absence or presence of 30 µM Yoda1. **b**, Scatter plot of the maximal poking-induced currents of the indicated mutants normalized to the mPiezo1 current. Each bar represents mean ± s.e.m. **c**, Yoda1-induced fold changes of the poking-induced Imax. Each bar represents mean ± s.e.m. **d**, Representative stretching-activated currents at -80 mV. **e**, Scatter plot of the maximal stretching-induced currents. Statistical significance was assessed using one-way ANOVA with Dunn's comparison to mPiezo1. Each bar represents mean ± s.e.m. $^{***}P < 0.0001$. **f**, Yoda1 dose response curves of the indicated constructs assayed by FLIPR. The curves were fitted with a Boltzmann equation. Each data point represents mean ± s.e.m., n = 4 wells. **g**, Relationship between poking-induced currents and the applied poking displacement. Each data point represents mean ± s.e.m. **h**, Scatter plot of the maximal poking-induced currents recorded at -60 mV. Each bar represents mean ± s.e.m. Statistical significance was assessed using the unpaired, two tailed Student's t-test. $^{***}P < 0.0001$. **i**, Scatter plot of the maximal stretching-induced currents. Each bar represents mean ± s.e.m. Statistical significance was assessed using the unpaired, two tailed Student's t-test. $^{**}P = 0.0045$. **j**, Pressure-current relationships of the stretching-induced currents. The curves were fitted with a Boltzmann equation. Each data point represents mean ± s.e.m. **k**, Scatter plot of the P_{50} calculated from the fit of the pressure-current relationship of individual recordings with a Boltzmann equation. Each bar represents mean ± s.e.m. Statistical significance was assessed using the unpaired, two tailed Student's t-test. $^{**}P = 0.0091$. The number of recorded cells is shown above the bar.

ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Molecular cloning. All constructs were subcloned by using the One Step Cloning Kit (Vazyme Biotech) according to the instruction manual and as described previously[3], then sequenced to validate the desired mutations. The loop deletion mutants include ΔL3-4 (84-122), ΔL7-8 (275-317), ΔL11-12 (492-521), ΔL15-16 (657-677), ΔL19-20 (870-921), and ΔL23-24 (1060-1150).

Protein expression and purification. The purification of the mPiezo1 protein was performed essentially as in our previously described protocols except for two modifications to the details[3]. First, the time for purification was shortened to within two days. Second, the collection volume of each peak fraction was reduced to obtain more homogeneous fractions. The fresh protein, not subjected to a concentration step, was then directly used for cryo-sample preparation.

Sample preparation and cryo-electron microscopy data acquisition. Four-microliter aliquots of detergent-solubilized mPiezo1 at a concentration of approximately 0.18 mg/ml were applied to glow-discharged 300-mesh Quantifoil R2/2 grids (Quantifoil, Micro Tools GmbH, Germany) coated with a homemade continuous thin carbon layer. After a 15 sec waiting time, the grids were blotted for 3.5 sec and plunged into liquid ethane using an FEI Mark IV Vitrobot operated at 8 °C and 100% humidity. The grids were transferred to a Titan Krios (FEI) electron microscope operating at a voltage of 300 kV with a K2 Summit direct electron detector (Gatan, Inc.) in the super-resolution counting mode. Data acquisition was performed using UCSF-Image4[30] with a nominal magnification of 22,500 X, which yields a super-resolution pixel size of 0.66 Å on image plane, and with defocus ranging from -1.5 μm to -3.0 μm. The dose rate on the detector was ~8.2 counts per pixel per second with a frame exposure time of 0.25 second and a total exposure time of 8 seconds. Each micrograph stack contains 32 frames. The total dose rate was approximately 50 e-/Å^2 for each micrograph.

Image processing. A simplified diagram of the procedure for image processing is presented in Extended Data Fig. 2. A total of 5840 and 4331 cryo-EM micrographs were manually collected for the mPiezo1 proteins. The motion correction was performed using MotionCor1[31] with 2×2 binning, resulting in a pixel size of 1.32 Å. The output stacks from MotionCor1 were further motion corrected with MotionCor2[32], and meanwhile, dose weighting was performed, yielding motion-corrected integrated images for further processing. The whole image defocus parameters were estimated by CTFFIND3[33]. A total of 1,417,116 and 1,405,683 particles were respectively autopicked from the 2 batches of the datasets by RELION 1.4[34]. Several rounds of two-dimensional (2D) classification were then performed. Those particles in the classes with fuzzy class averages were considered as bad particles and excluded from further analysis. Finally, 377,175 and 203,621 particles from the two batches were selected for further three-dimensional (3D) analysis, respectively. A cylinder model generated by SPIDER[35] was used as the initial model for the first round of 3D refinement. Then, each particle was recentered using the in-plane translations measured in 3D refinement and re-extracted from the motion-corrected integrated micrographs. Gctf[36] was used to refine the local defocus parameters. The well centered particles with more accurate defocus parameters were subjected to further 3D refinement, which resulted in two electron density maps at 4.25 Å and 4.86 Å resolution, respectively. The particles from the two datasets were pooled together to expand the data volume. A total of 465,503 particles with clear secondary-structure features in the corresponding class averages were selected after further 2D classification, and subjected to 3D refinement, resulting in a map at 4.15 Å resolution. To further eliminate heterogeneous particles, we used a random-phase 3D classification method[37], and the remained 238,529 good particles were subjected to a final 3D refinement. Eventually, the resolution was improved to 3.97 Å.

The particles selected from the random-phase 3D classification were also subjected to 3D classification with C3 symmetry using refined orientations based on the 3.97 Å map, while no image alignment was performed in the 3D classification to limit the uncertainty of the angular search. The particles of the two classes with the largest variations were chosen and subjected to one additional round of 3D auto-refinement, which resulted in two reconstructions with overall resolutions of 4.51 Å and 4.40 Å. To examine whether the three subunits within the trimeric channel might move independently, the 4.15 Å reconstruction was subjected to an additional round of 3D classification without symmetry.

To improve the density of the far end of the blade region, another two replicas of particles with -120°/120° rotation on the pseudo-C3 axis were added to the original particles, and then projections of the cap and the other two blades were subtracted from the tripled experimental particle images. The remaining particles were processed following a focused classification procedure[38], which resulted in a blade map at 4.86 Å resolution showing better resolved TMs. To improve the density of the linker between the termini of the CED and the OH and IH projections of the peripheral regions of the blades were subtracted from the experimental images, and the remaining particles were processed following a focused classification procedure, which resulted in a pore map at 4.25 Å resolution.

The reported resolutions are based on the gold-standard Fourier shell correlation 0.143 criterion[39]. All density maps were sharpened by applying a negative B-factor that was estimated using automated procedures[40]. Local resolution variations were estimated using Blocres[41].

Model building and structure refinement. The TM regions of mPiezo1 in the sequence were predicted by HMMTOP[42]. The crystal structure of CED (PDB ID: 4RAX[3]) was docked into the cryo-EM map in Chimera[43]. The remaining part of the model was built *de novo* in COOT[44]. The cryo-EM map showed clear side chain features in the C-terminal region, which allowed us to build the atomic model including side chains from residues 972 to the C-terminal 2547, with the remaining TMs as poly-alanines. The topology of the whole structure was refined using MDFF[45] and phenix.real_space_refine[46] with NCS restraints.

Piezo1 cross-linking and LC-MS/MS analysis. The purified Piezo1 proteins were cross-linked at room temperature with DSS (disuccinimidyl suberate), BS3 (bis[sulfosuccinimidyl] suberate), and sulfo-GMBS (N-[g-maleimidobutyryloxy] sulfosuccinimide ester), whose spacer arm lengths are 7.3 Å, 11.4 Å and 11.4 Å, respectively. The samples were then digested and subjected to LC-MS/MS analysis. The cross-linked peptides were identified using the pLink software as previously described[47].

Immunostaining. Live-cell labeling was carried out following the exact procedure reported previously[4].

Whole-cell electrophysiology and mechanical stimulation. The protocols for HEK293T cell culture, transient transfection and patch-clamp experiments with an Axopatch 200B amplifier (Axon Instruments) or HEKA EPC10 were essentially similar to those previously described[4]. For whole-cell patch-clamp recordings, the recording electrodes had a resistance of 2-3 MΩ when filled with an internal solution composed of (in mM) 133 CsCl, 1 $CaCl_2$, 1 $MgCl_2$, 5 EGTA, 10 HEPES (pH 7.3 with CsOH), 4 MgATP and 0.4 Na_2GTP. The extracellular solution was composed of (in mM) 133 NaCl, 3 KCl, 2.5 $CaCl_2$, 1 $MgCl_2$, 10 HEPES (pH 7.3 with NaOH) and 10 glucose. All experiments were carried out at room temperature. The currents were sampled at 20 kHz, filtered at 2 kHz using the Clampex 10.4 software (Axon Instruments) or Patchmaster software. Leak currents before mechanical stimulation were subtracted off-line from the current traces.

Mechanical stimulation was delivered to the cell during the patch-clamp recording at an angle of 80° using a fire-polished glass pipette (tip diameter 3 – 4 μm) as previously described[1,2]. The downward movement of the probe toward the cell was driven by a Clampex controlled piezo-electric crystal micro-stage (E625 LVPZT Controller/Amplifier; Physik Instrument). The probe had a velocity of 1 μm/ms during the downward and upward motion, and the stimulus was maintained for 150 ms. A series of mechanical steps in 1 μm increments was applied every 20 s, and the currents were recorded at a holding potential of -70 mV. Yoda1 was solubilized in DMSO as a stock solution of 30 mM and diluted to a final concentration of 30 μM using the extracellular solution. The poking-induced currents were recorded within 10 min using the extracellular solution with or without 30 μM Yoda1.

Cell-attached electrophysiology. Stretching-activated currents were recorded in the cell-attached patch-clamp configuration as previously described[4]. The currents were sampled at 20 kHz and filtered at 2 kHz. The pipettes were filled with a solution consisting of (in mM) 130 NaCl, 5 KCl, 10 HEPES, 1 $CaCl_2$, 1 $MgCl_2$, and 10 TEA-Cl (pH 7.3, balanced with NaOH). The external solution used to zero the membrane potential consisted of (in mM) 140 KCl, 10 HEPES, 1 $MgCl_2$, and 10 glucose (pH 7.3 with KOH). All experiments were performed at room temperature. The membrane patches were stimulated with negative pressure pulses for 500 ms through the recording electrode using a Patchmaster controlled pressure clamp HSPC-1 device (ALA-scientific). Stretching-activated channels were recorded at a holding potential of -80 mV with pressure steps from 0 to -100 mm Hg (-10 mm Hg increments), and 4-11 recording traces were averaged per cell for analysis. The current-pressure relationships were fitted with a Boltzmann equation of the form $I(P) = [1 + \exp(-(P - P_{50})/s)]^{-1}$, where I is the peak of the stretching-activated current at a given pressure, P is the applied patch pressure (in mm Hg), P_{50} is the pressure value that evoked a current value of 50% of Imax, and s reflects the current sensitivity to pressure.

Fluorescent imaging plate reader (FLIPR). Human embryonic kidney 293T (HEK293T) cells were grown in Dulbecco's Modified Eagle Medium containing 4.5 mg/ml glucose, 10% fetal bovine serum, 1% penicillin/streptomycin. Cells were seeded in 50 μg/ml poly-D-lysine-coated 96-well plates (3 x 10^4 cells/well) and allowed to grow for approximately 18 h, then co-transfected with a total of 250 ng of cDNA containing Piezo1 or the mutants and the genetically encoded Ca^{2+} indicator, GCAMP6s, using Lipofectamine 2000 (Invitrogen, Life Technology). Two days after transfection, the cells were washed with buffer containing 1xHBSS (1.3 mM Ca^{2+}) and 10 mM HEPES (pH 7.2). The 96-well cell plate with 50 μl of buffer in each well was then transferred to the FLIPR Tetra (Molecular Device) for recording the GCAMP6s fluorescent signal at a time interval of 1s using an excitation wavelength of 470-490 nm and an emission wavelength of 515-575 nm. The baseline

ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

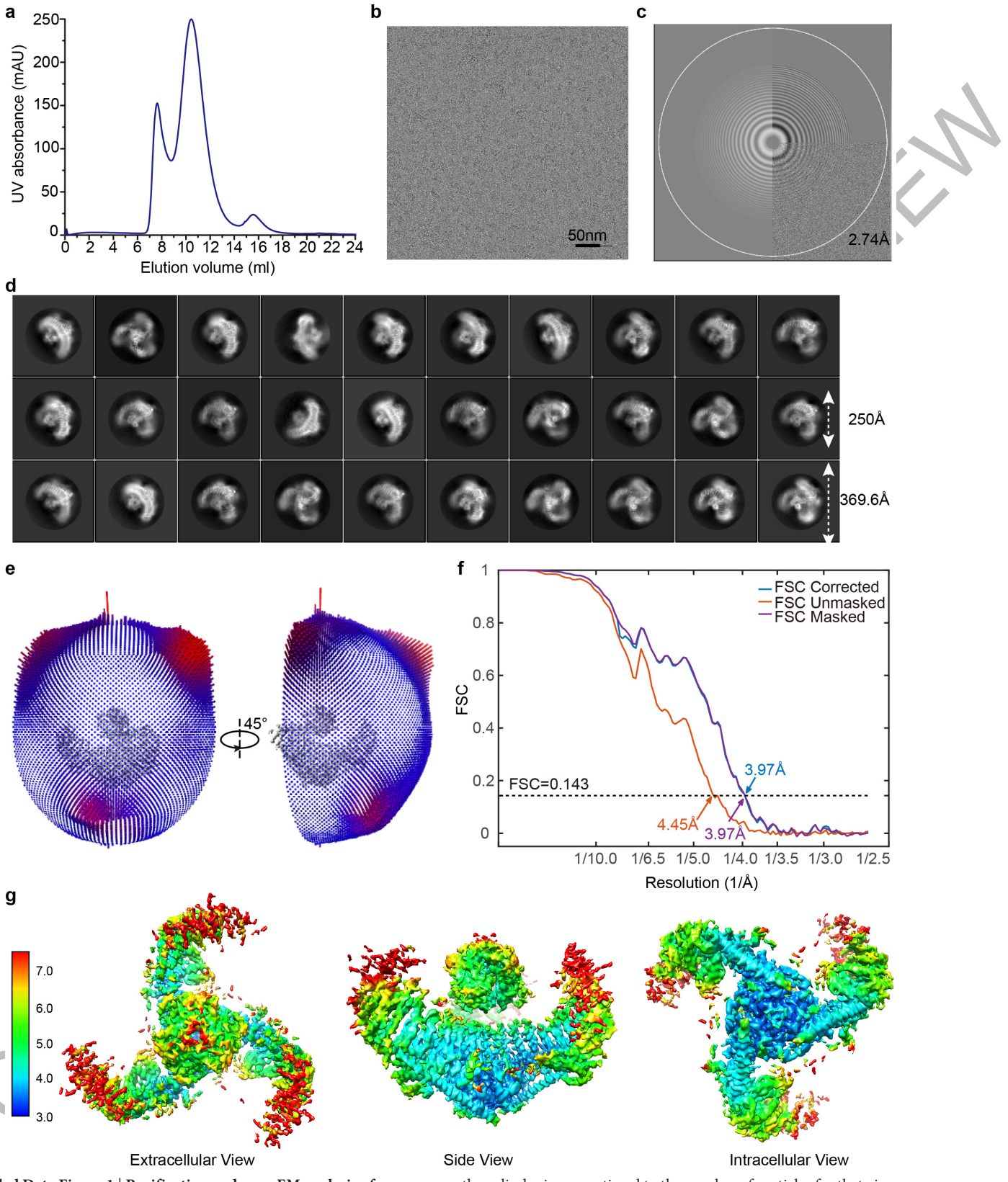
GCAMP6s fluorescent signal without Yoda1 addition was monitored for 15 s, and then the response to Yoda1 was recorded for 200 s after transferring 50 μl Yoda1 at 2 times of the final concentrations from a separate compound plate into the cell plate. The Yoda1-induced fluorescent signal change was calculated by subtracting the baseline fluorescent signal from the peak fluorescent signal. For each dose point of Yoda1, the data were averaged from 4 repetitive wells. The representative traces of the fluorescence signal change of the GCAMP6s in response to 30 μM Yoda1 are shown in Extended Data Figure 8i.

Data Availability. The structural coordinates of mPiezo1 were deposited in the Protein Data Bank under the ID 5Z10. The Cryo-EM map was deposited into the Electron Microscopy Data Bank (EMDB) under the accession ID 6865.

30. Li, X. M., Zheng, S., Agard, D. A. & Cheng, Y. F. Asynchronous data acquisition and on-the-fly analysis of dose fractionated cryoEM images by UCSFImage. *J. Struct. Biol.* **192**, 174–178, doi:10.1016/j.jsb.2015.09.003 (2015).
31. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature methods* **10**, 584–590, doi:10.1038/nmeth.2472 (2013).
32. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature methods* **14**, 331–332, doi:10.1038/nmeth.4193 (2017).
33. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
34. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology* **180**, 519–530, doi:10.1016/j.jsb.2012.09.006 (2012).
35. Shaikh, T. R. *et al.* SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nature protocols* **3**, 1941–1974, doi:10.1038/nprot.2008.156 (2008).
36. Zhang, K. Gctf & Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12, doi:10.1016/j.jsb.2015.11.003 (2016).
37. Gong, X. *et al.* Structural Insights into the Niemann-Pick C1 (NPC1)-Mediated Cholesterol Transfer and Ebola Infection. *Cell* **165**, 1467–1478, doi:10.1016/j.cell.2016.05.022 (2016).
38. Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human gamma-secretase. *eLife* **4**, doi:10.7554/eLife.11182 (2015).
39. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature methods* **9**, 853–854, doi:10.1038/nmeth.2115 (2012).
40. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
41. Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *Journal of structural biology* **184**, 226–236, doi:10.1016/j.jsb.2013.08.002 (2013).
42. Tusnady, G. E. & Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics (Oxford, England)* **17**, 849–850 (2001).
43. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605–1612, doi:10.1002/jcc.20084 (2004).
44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486–501, doi:10.1107/s0907444910007493 (2010).
45. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure (London, England : 1993)* **16**, 673–683, doi:10.1016/j.str.2008.03.005 (2008).
46. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213–221, doi:10.1107/s0907444909052925 (2010).
47. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nature methods* **9**, 904–906, doi:10.1038/nmeth.2099 (2012).

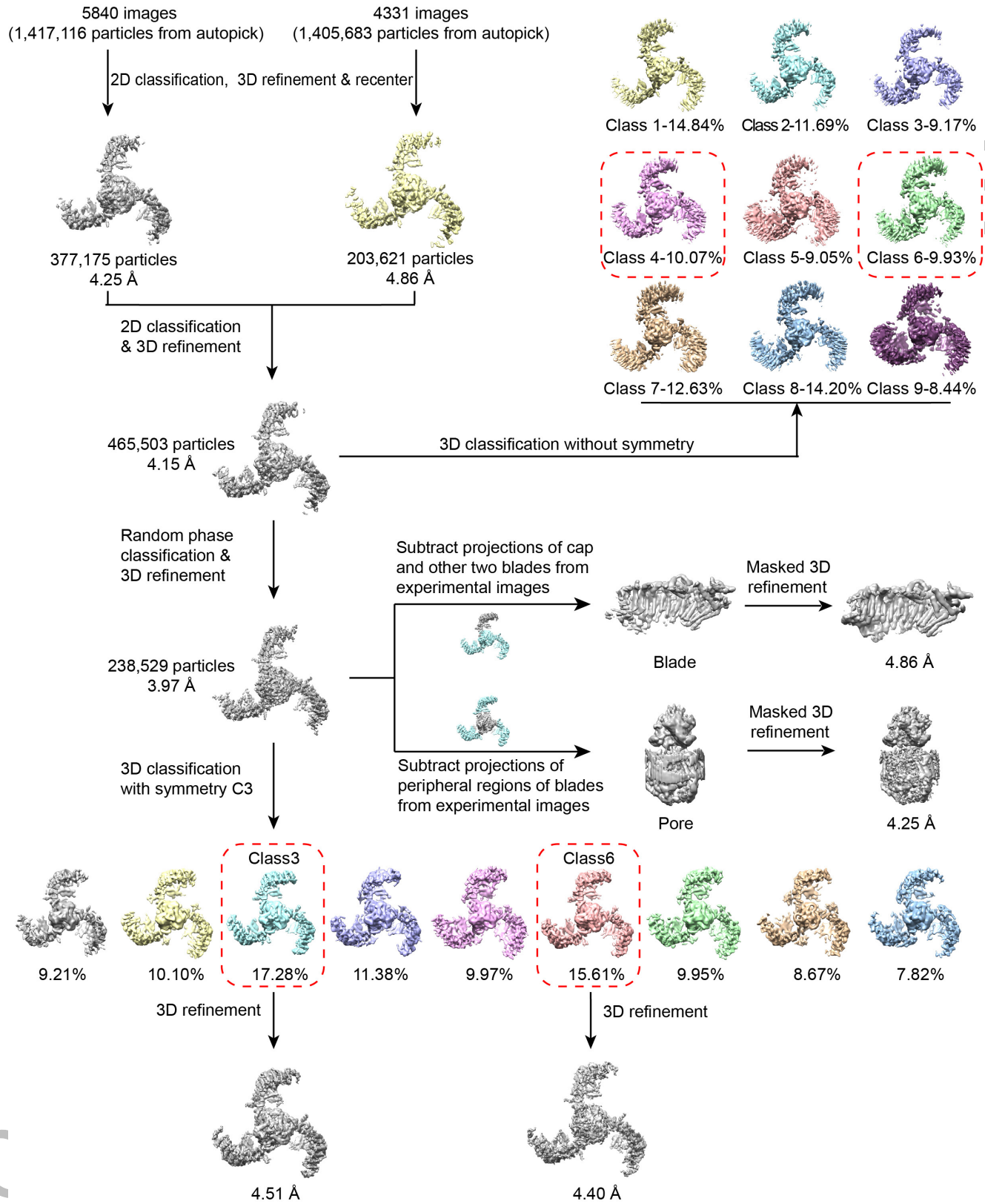
ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



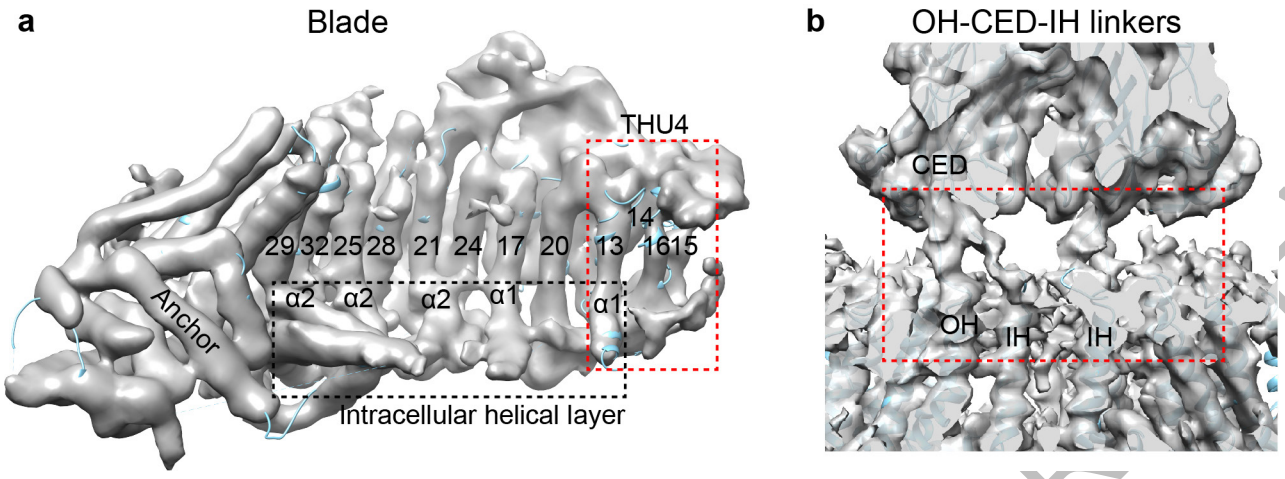
Extended Data Figure 1 | Purification and cyro-EM analysis of mPiezo1. a, A representative trace of gel filtration of the full-length mPiezo1. UV, ultraviolet. The experiment was independently repeated for more than three times with similar results. **b**, A representative cryo-electron micrograph of mPiezo1. The experiment was independently repeated for more than three times with similar results. **c**, Power spectrum of the micrograph in **b**, with the 2.74Å frequency indicated. **d**, Representative 2D class averages of mPiezo1 particles. **e**, Euler angle distribution of particles used in the final 3D reconstruction, the height of the cylinder is proportional to the number of particles for that view. **f**, Gold-standard Fourier shell correlation (FSC) curves of the final density map. The FSC curves were calculated with (purple) or without (red) the application of a soft mask to the two half-set maps. The final FSC curve (blue) was corrected for the soft-mask-induced effect. Reported resolutions were based on the FSC = 0.143 criteria. **g**, The final 3D density map of mPiezo1 shown in the indicated views is coloured according to the local resolutions estimated by the software of Blocres.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 2 | Flowchart of EM data processing. Details of data processing were described in the Imaging processing part of Methods.

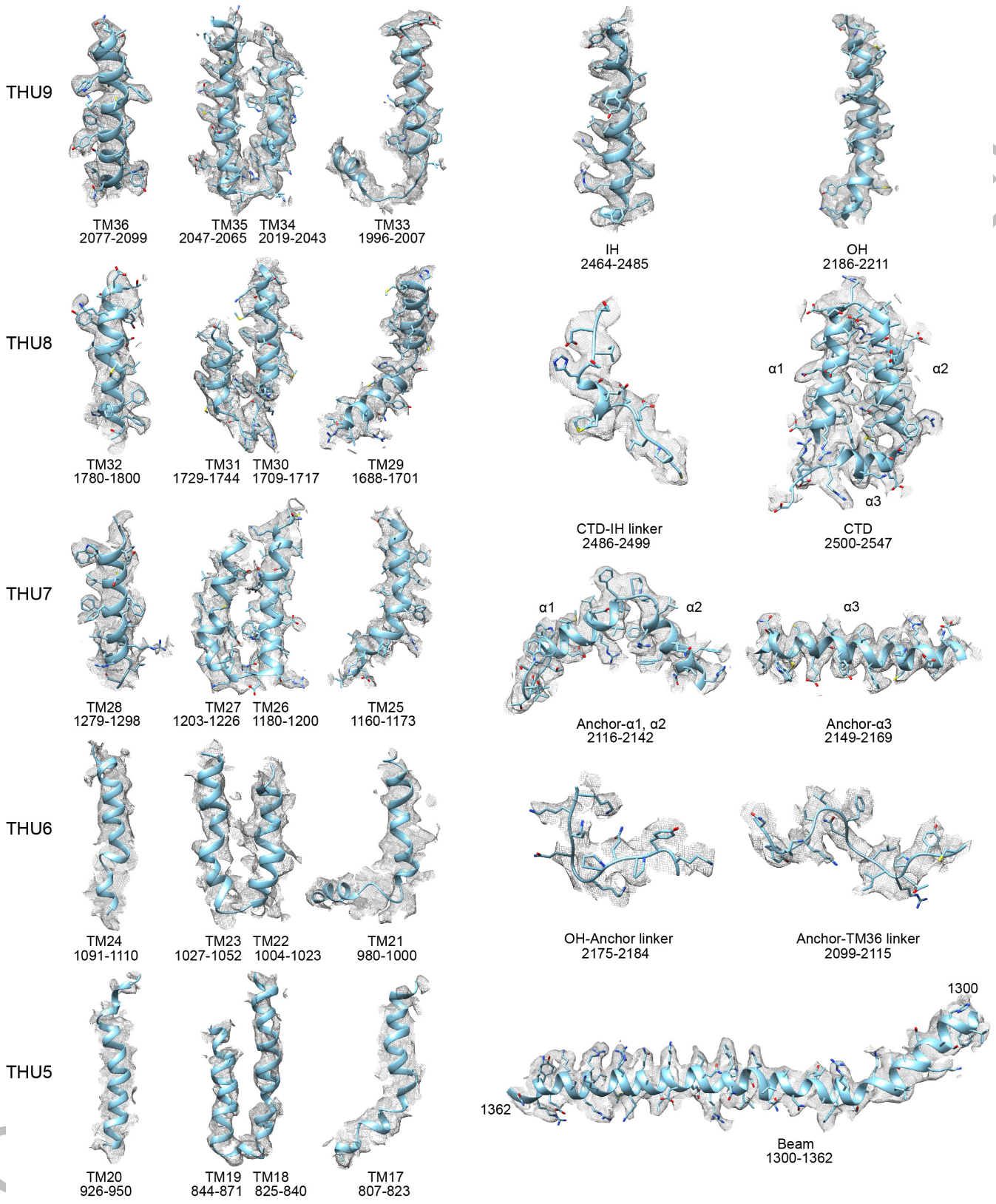
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 3 | Subtraction of the projection. a, Subtraction of the projection of the cap and the other two blades. A distinguishable map of THU4, comprising TM13-16 is shown in the red dashed box. Intracellular helical layer, containing several α-helixes respectively connecting to TM29, TM25, TM21, TM17, TM13, is highlighted in the black dashed box. **b**, Subtraction of the projection of the three blades projection. Identifiable linkers between OH and CED as well as IH and CED are shown in the red dashed box.

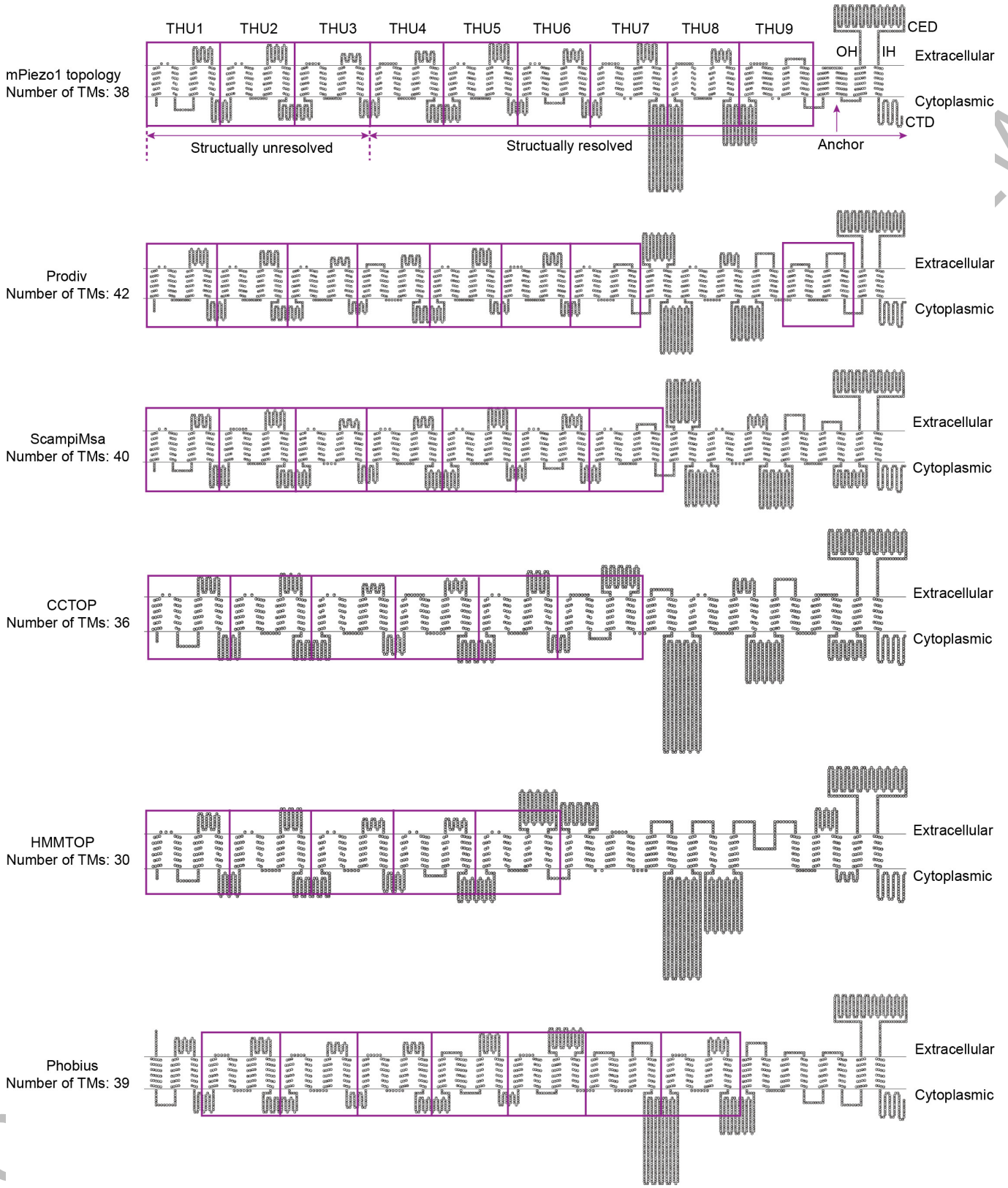
ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



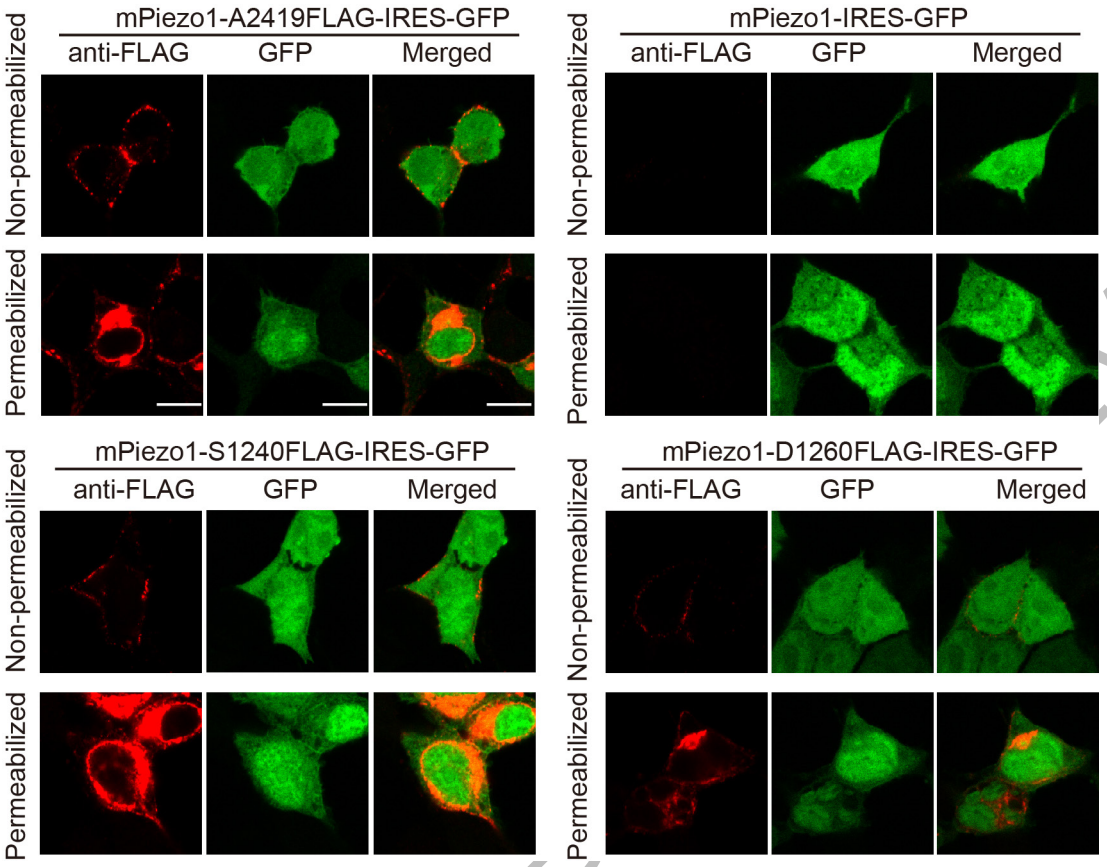
Extended Data Figure 4 | View of the indicated structural domains illustrates the quality of the cryo-EM density of mPiezo1. The helices are shown in cartoon representation with side chains as sticks. The cryo-EM density is shown as gray mesh.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 5 | Membrane topology of mouse Piezo1. Based on various membrane topology prediction algorithms, the number of TMs for mPiezo1 varies from 30 to 42. The N-terminal region contains unanimously predicted THUs (highlighted in the purple boxes) showing typical features of the structurally revealed THU7-9. On the basis the resolved 3D structure and the predicted THUs, we propose that mPiezo1 possesses a 38-TM topology comprising 9 tandem THUs and the OH and IH. Diagrams were drawn using the TOPO2 program.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 6 | The TM27-28 loop containing the S1240 and D1260 residues is located at the extracellular side. Immunofluorescent staining images of cells transfected with the indicated constructs using the anti-FLAG antibody either in live-labeling (upper panels) or after fixation and permeabilization (lower panels). Scale bars, 10 μm. GFP, green fluorescent protein; IRES, internal ribozyme entry site. The experiments were repeated in two coverslips with similar results.

ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

a

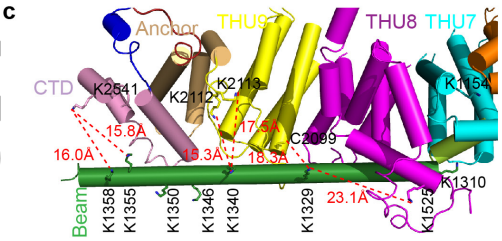
BS^3/DSS (K-K)		Site 1	-	Site2			#Spec-Total	Best E-value
THU8-9	TH32-33	1815	-	145	TH4-5	THU1-2	2	4.9E-05
			-	1828	TH32-33	THU8-9	10	1.5E-08
			-	1879			9	9.7E-08
		1828	-	1815	TH32-33	THU8-9	3	1.2E-05
			-	1823			76	2.2E-09
			-	1862			8	3.8E-08
			-	1879			3	2.7E-05
			-	1969			8	5.0E-07
		1834	-	1828	TH32-33	THU8-9	9	1.1E-10
		1847	-	349	TH8-9	THU2-3	1	1.6E-05
			-	1372	TH28-29	THU7-8	3	1.8E-05
			-	1815	TH32-33	THU8-9	17	6.2E-11
			-	1828			5	8.4E-06
			-	1834			20	2.6E-07
			-	1862			157	2.0E-14
			-	1868			48	4.1E-12
			-	1879			19	7.9E-07
			-	1894			8	5.3E-10
			-	1921			1	3.0E-05
			-	1937			4	1.4E-08
			-	1969			1	3.7E-06
		1862	-	414	TH8-9	THU2-3	10	8.6E-08
			-	1372	TH28-29	THU7-8	7	1.3E-08
			-	1815	TH32-33	THU8-9	17	2.3E-09
			-	1823			4	3.9E-06
			-	1828			19	3.5E-07
			-	1879			16	4.9E-07
			-	1894			11	2.2E-12
			-	1921			13	1.1E-06
			-	1929			7	3.2E-09
			-	1937			16	1.4E-07
			-	1969			15	5.3E-11
		1868	-	1862	TH32-33	THU8-9	89	1.5E-12
			-	1894			1	3.1E-05
			-	1921			1	6.9E-06
			-	1937			2	8.3E-06
		1879	-	414	TH8-9	THU2-3	1	4.5E-08
			-	1862	TH32-33	THU8-9	9	6.6E-06
			-	1894			34	1.6E-09
			-	1937			1	7.9E-05
		1894	-	1815	TH32-33	THU8-9	3	1.8E-07
			-	1823			1	5.6E-05
			-	1879			9	2.2E-11
			-	1921			10	2.7E-07
			-	1929			1	4.0E-06
			-	1937			6	3.5E-07
		1915	-	145	TH4-5	THU1-2	1	1.7E-05
			-	1862	TH32-33	THU8-9	1	1.5E-05
			-	1879			13	4.4E-07
			-	1894			35	2.9E-10
			-	1900			1	7.1E-05
			-	1921			17	1.5E-06
			-	1937			4	4.7E-07
		1929	-	145	TH4-5	THU1-2	1	3.8E-05
			-	1921	TH32-33	THU8-9	81	1.2E-09
			-	1937			89	4.1E-15
		1937	-	414	TH8-9	THU2-3	8	1.1E-09
			-	1879	TH32-33	THU8-9	13	4.0E-07
			-	1921			39	3.8E-10
			-	1931			39	1.6E-06
		1969	-	414	TH8-9	THU2-3	6	5.3E-08
			-	1815	TH32-33	THU8-9	27	2.0E-12
			-	1879			13	7.9E-08
			-	1937			3	8.5E-06

BS^3/DSS (K-K)		Site 1	-	Site2			#Spec-Total	Best E-value
Beam		1310	-	145	TH4-5	THU1-2	1	1.7E-06
			-	775	TH16-17	THU4-5	2	4.2E-07
			-	1525	TH28-29	THU7-8	10	2.9E-10
			-	1862	TH32-33	THU8-9	7	1.4E-16
			-	1879			3	1.5E-06
			-	1969			10	3.1E-11
		1329	-	1154	TH24-25	THU6-7	1	3.1E-05
			-	1310	TH28-29	THU7-8	19	5.2E-11
			-	1340	TH32-33	THU8-9	9	1.2E-09
			-	1525			32	3.0E-10
			-	1815			6	4.4E-08
			-	1862			16	7.3E-09
			-	1879			1	8.7E-05
			-	1894			1	5.1E-06
			-	1969			6	1.5E-07
		1340	-	1525	TH28-29	THU7-8	4	9.1E-06
			-	1823	TH32-33	THU8-9	2	4.6E-05
			-	1862			6	3.8E-09
			-	2112	TH36-37		25	5.6E-09
			-	2113			6	5.2E-09
		1346	-	186	TH4-5	THU1-2	1	6.0E-06
			-	349	TH8-9	THU2-3	6	1.5E-07
			-	1350	TH28-29	THU7-8	16	4.6E-06
			-	1815	TH32-33	THU8-9	3	5.6E-06
			-	1823			2	1.5E-05
			-	1828			1	7.4E-05
			-	1862			7	1.1E-05
			-	1879			17	8.0E-07
			-	1937			1	5.0E-06
			-	1969			12	3.7E-08
			-	2541	TH38-C	CTD	1	2.9E-07
		1350	-	1879	TH32-33	THU8-9	7	6.1E-08
		1355	-	1847	TH32-33	THU8-9	1	4.2E-05
			-	1862			1	1.7E-05
			-	2541	TH38-C	CTD	2	4.0E-05
		1358	-	1350	TH28-29	THU7-8	5	7.4E-06
			-	1372			8	3.3E-09
			-	1862	TH32-33	THU8-9	3	7.0E-10
			-	2541	TH38-C	CTD	1	2.6E-06
THU1-2	TH4-5	145	-	414	TH8-9	THU2-3	11	2.1E-08
			-	1921	TH32-33	THU8-9	3	3.8E-06
			-	1937			3	2.3E-05
		178	-	145	TH4-5	THU1-2	7	5.3E-08
			-	414	TH8-9	THU2-3	8	9.8E-06
			-	1815	TH32-33	THU8-9	1	6.0E-06
			-	1937			10	5.1E-09
		186	-	145	TH4-5	THU1-2	10	1.5E-05
			-	349	TH8-9	THU2-3	7	2.3E-08
			-	414			2	2.5E-05
			-	1879	TH32-33	THU8-9	11	4.9E-09
			-	1921			6	2.0E-08
			-	1937			18	1.0E-09
THU2-3	TH8-9	349	-	145	TH4-5	THU1-2	4	9.0E-06
			-	414	TH8-9	THU2-3	3	9.6E-10
			-	1937	TH32-33	THU8-9	4	2.4E-07
THU4-5	TH16-17	775	-	145	TH4-5	THU1-2	19	4.6E-13
			-	186			17	1.2E-10
			-	349	TH8-9	THU2-3	4	2.4E-09
			-	414			5	1.9E-06
			-	1862	TH32-33	THU8-9	12	4.8E-07
			-	1937			9	7.2E-09
THU7-8	TH28-29	1372	-	1879	TH32-33	THU8-9	3	9.6E-06
		1525	-	1778			1	3.5E-05
	TH36-37	2112	-	2174	TH36-37		6	3.2E-07
		2113	-	2174	TH36-37		44	1.3E-11
CED	TH37-38	2374	-	2388	TH37-38	CED	38	4.7E-16
			-	2438			2	4.5E-06
		2388	-	2329	TH37-38	CED	3	1.4E-06
			-	2374			43	3.7E-16
CTD	TH38-C	2505	-	2112	TH36-37		2	1.2E-08
		2541	-	1372	TH28-29	THU7-8	8	2.6E-10
			-	1862	TH32-33	THU8-9	5	5.1E-14
			-	1879			4	7.6E-08

b

sulfo-GMBS (K-C)		Site 1	-	Site2			#Spec-Total	Best E-value
THU2-3	TH8-9	349	-	343	TH8-9	THU2-3	1	3.9E-06
THU7-8	Beam	1329	-	2099	TH36-37		65	5.5E-10
	TH28-29	1525	-	2099			3	3.9E-05
THU8-9	TH32-33	1823	-	2099			6	5.6E-10
		1862	-	2099			12	5.3E-07
		1969	-	2099			128	3.4E-13
		1969	-	1818	TH32-33	THU8-9	8	1.7E-08
	TH36-37	2112	-	2099	TH36-37		24	1.5E-06

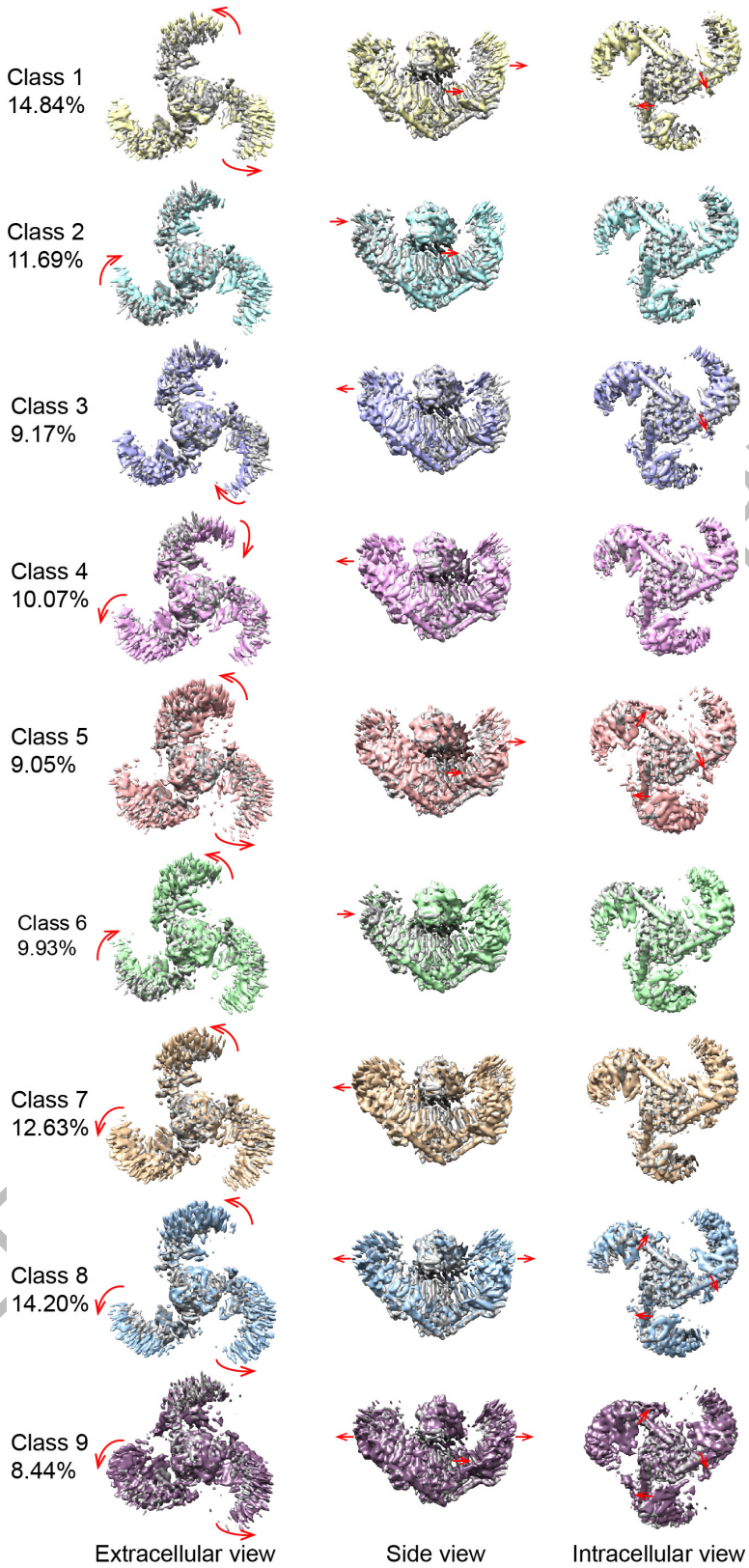
c



Extended Data Figure 7 | Chemically crosslinked lysine-lysine and lysine-cysteine pairs identified in mPiezo1. a and **b**, Purified mPiezo1 proteins were crosslinked with BS^3/DSS (**a**) or sulfo-GMBS (**b**), and then digested with trypsin. Following LC–MS/MS analysis of the peptides, crosslinked lysine pairs were identified using pLink. **c**, The diagram shows the crosslinked lysine pairs between residues located in the beam and other regions. Atom-to-atom distance of the crosslinked residues is shown.

ACCEPTED MANUSCRIPT VIEW

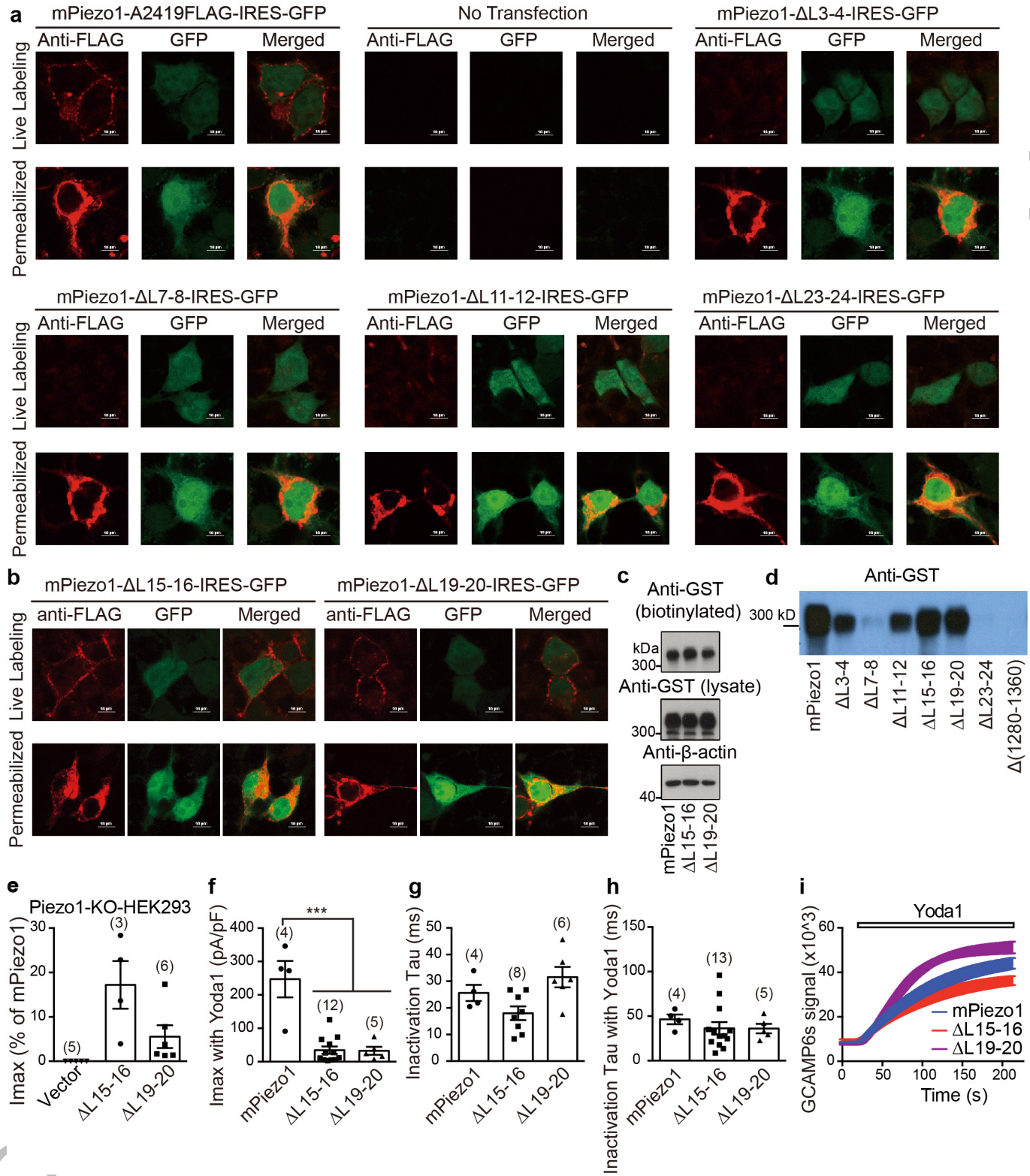
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | Conformational heterogeneity of Piezo1. 9 classes of Piezo1 structures resulting from symmetry-free 3D classification. Conformational heterogeneity is shown by comparing different classes with the low-passed 6 Å map of the 3.97 Å map with C3 symmetry. Red arrows represent the relative movements of the blades and beams.

ACCELERATED PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



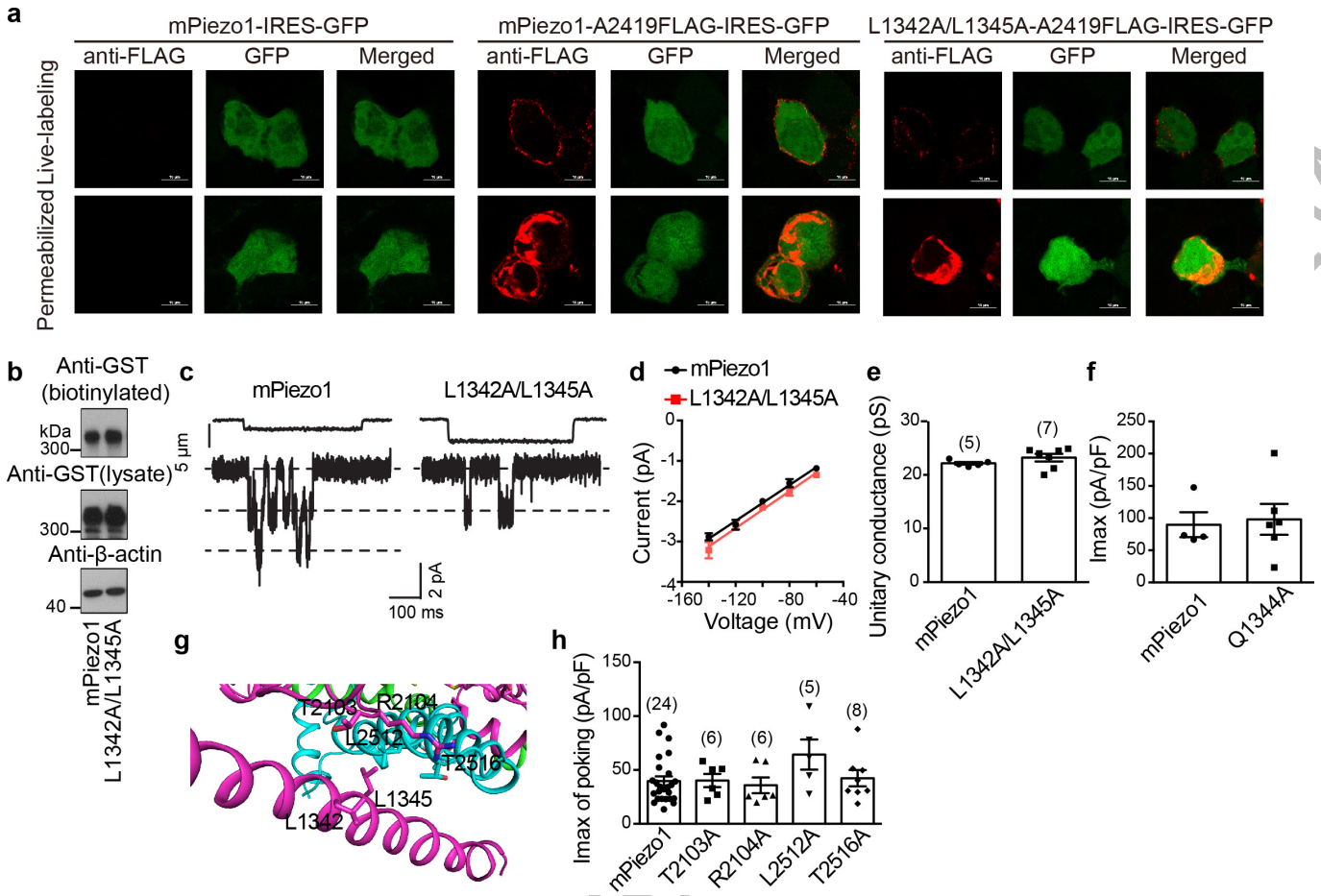
Extended Data Figure 9 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 9 | Characterization of the Piezo1 deletion mutants. a - **b**, Immunofluorescent staining images of cells transfected with the indicated constructs using the anti-FLAG antibody either in live-labeling (upper panels) or after fixation and permeabilization (lower panels). Scale bars, 10 μm. GFP, green fluorescent protein; IRES, internal ribozyme entry site. The experiments were repeated in two coverslips with similar results. **c**, Cell surface biotinylation assay showing comparable plasma membrane expression of the indicated constructs. The experiment was independently repeated for two times with similar results. **d**, The GST-tagged proteins were pulled-down by glutathione beads, followed with western blotting using the anti-GST antibody. The experiment was independently repeated for two times with similar results. **e**, Scatter plot of the maximal poking-induced currents of Piezo1-KO-HEK293 cells transfected with the indicated mutants, which were normalized to the mPiezo1 current. Each bar represents mean ± s.e.m., and the number of recorded cells are shown above the bar. **f**, Scatter plot of the maximal poking-induced currents of HEK293T cells transfected with the indicated constructs in the presence of 30 μM Yoda1. Each bar represents mean ± s.e.m., and the number of recorded cells are shown above the bar. Statistical significance was assessed using one-way ANOVA with Dunn's comparison to mPiezo1. $P < 0.0001$. **g** and **h**, Scatter plot of the inactivation tau of HEK293T cells transfected with the indicated constructs in the absence (**g**) or presence (**h**) of 30 μM Yoda1. Each bar represents the mean ± s.e.m., and the number of recorded cells are shown above the bar. Statistical significance was assessed using one-way ANOVA with Dunn's comparison to mPiezo1. The numbers of recorded cells are shown above the bar. In **h**, the P value for ΔL15-16 and ΔL19-20 in the absence of Yoda1 (**g**) is 0.2267 and 0.4177, and in the presence of Yoda1 (**h**) is 0.6263 and 0.6934, respectively. **i**, Representative traces of the fluorescence signal change of the genetically encoded Ca^{2+} indicator, GCAMP6s, from HEK293T cells co-transfected with the indicated constructs and GCAMP6s, in response to 30 μM Yoda1. The experiment was independently repeated for three times with similar results.

ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | Characterizations of the L1342A/L1345A mutant. a, Immunofluorescent staining images with the anti-FLAG antibody either in live-labeling (upper panels) or after fixation and permeabilization (lower panels). The experiment was independently repeated for three times with similar results. **b**, Cell surface biotinylation assay showing comparable plasma membrane expression of the indicated constructs. The experiment was independently repeated for two times with similar results. **c**, Representative single-channel current traces of the indicated constructs recorded at -140 mV. **d**, Linear regression fit of average I-V relationships of single-channel recordings of the indicated constructs. Each data point represents the mean ± s.e.m. The number of recorded cells for mPiezo1 and L1342A/L1345A is 5 and 7, respectively. **e**, Scatter plot of the unitary conductance calculated from fit of individual recordings. Each bar represents the mean ± s.e.m., and the number of recorded cells are shown above the bar. Statistical significance was assessed using the unpaired, two tailed Student's t-test. P = 0.2541. **f**, Scatter plot of the Imax of poking-induced currents. Each bar represents the mean ± s.e.m., and the number of recorded cells are shown above the bar. Statistical significance was assessed using the unpaired, two tailed Student's t-test. P = 0.8076. **g**, Structural representation of the L1342 and L1345 residues and those residues in close proximity. **h**, Scatter plot of the Imax of poking-induced currents of cells transfected with the indicated mutants. Statistical significance was assessed using One-way ANOVA with Dunn's comparison to mPiezo1. The P value for T2103A, R2104A, L2512A, T2516A is 0.9976, 0.9281, 0.3804 and 0.9997, respectively.

ACCELERATED ARTICLE PREVIEW

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature25436

A clumpy and anisotropic galaxy halo at redshift 1 from gravitational-arc tomography

Sebastian Lopez[1], Nicolas Tejos[2], Cédric Ledoux[3], L. Felipe Barrientos[4], Keren Sharon[5], Jane R. Rigby[6], Michael D. Gladders[7,8], Matthew B. Bayliss[9] & Ismael Pessa[4]

Every star-forming galaxy has a halo of metal-enriched gas that extends out to at least 100 kiloparsecs[1–3], as revealed by the absorption lines that this gas imprints on the spectra of background quasars[4]. However, quasars are sparse and typically probe only one narrow beam of emission through the intervening galaxy. Close quasar pairs[5–7] and gravitationally lensed quasars[8–11] have been used to circumvent this inherently one-dimensional technique, but these objects are rare and the structure of the circumgalactic medium remains poorly constrained. As a result, our understanding of the physical processes that drive the recycling of baryons across the lifetime of a galaxy is limited[12,13]. Here we report integral-field (tomographic) spectroscopy of an extended background source—a bright, giant gravitational arc. We can thus coherently map the spatial and kinematic distribution of Mg II absorption—a standard tracer of enriched gas—in an intervening galaxy system at redshift 0.98 (around 8 billion years ago). Our gravitational-arc tomography unveils a clumpy medium in which the absorption strength decreases with increasing distance from the galaxy system, in good agreement with results for quasars. Furthermore, we find strong evidence that the gas is not distributed isotropically. Interestingly, we detect little kinematic variation over a projected area of approximately 600 square kiloparsecs, with all line-of-sight velocities confined to within a few tens of kilometres per second of each other. These results suggest that the detected absorption originates from entrained recycled material, rather than in a galactic outflow.

We use the Multi Unit Spectroscopic Explorer (MUSE)[14] mounted on the European Southern Observatory Very Large Telescope to observe the 38″-long gravitational arc RCSGA 032727−132623 (ref. 15). This arc results from a lensed galaxy at redshift 1.70, which is highly magnified and stretched by a massive galaxy cluster at redshift 0.56 (Fig. 1). With a *g*-band magnitude of 19.15, it is among the brightest known arcs[16] and has a high surface brightness across a large area on the sky[17]. Magellan/MagE spectroscopy[18] of its brightest knot reveals the presence of a strong Mg II absorption system at redshift $z_{abs} = 0.98$, and we set out to map this absorption along the entire arc.

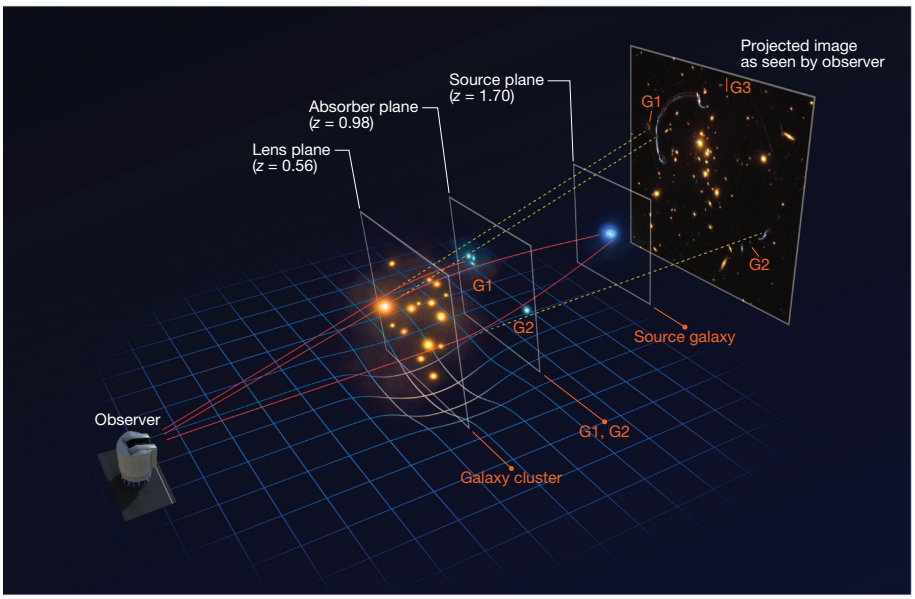


Figure 1 | Illustration of the lens geometry of the arc and the absorber in RCSGA 032727−132623. Light from the $z = 1.70$ background source galaxy ('source plane') is deflected and magnified (red lines) by an intervening galaxy cluster at $z = 0.56$ (ref. 15; 'lens plane') to form the bright giant arc that is seen projected (yellow dashed lines) onto the rightmost panel ('image plane'). As the light crosses the $z_{abs} = 0.98$ 'absorber plane', some of it is absorbed by Mg II in the gas that surrounds an absorbing galaxy that lies close to the arc in projection. Three candidates for such galaxies are detected at this redshift, marked 'G1', 'G2' and 'G3' in the image plane. This work deals with G1 and G2 (also indicated in the absorber plane), which probe the closest projected distances to the arc. Image produced by Carlos Polanco.

[1]Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile. [2]Instituto de Física, Pontificia Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile. [3]European Southern Observatory, Alonso de Córdova 3107, Casilla 19001, Vitacura 19, Santiago, Chile. [4]Instituto de Astrofísica, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile. [5]Department of Astronomy, University of Michigan, 500 Church Street, Ann Arbor, Michigan 48109, USA. [6]Observational Cosmology Laboratory, NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, Maryland 20771, USA. [7]Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA. [8]Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA. [9]Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

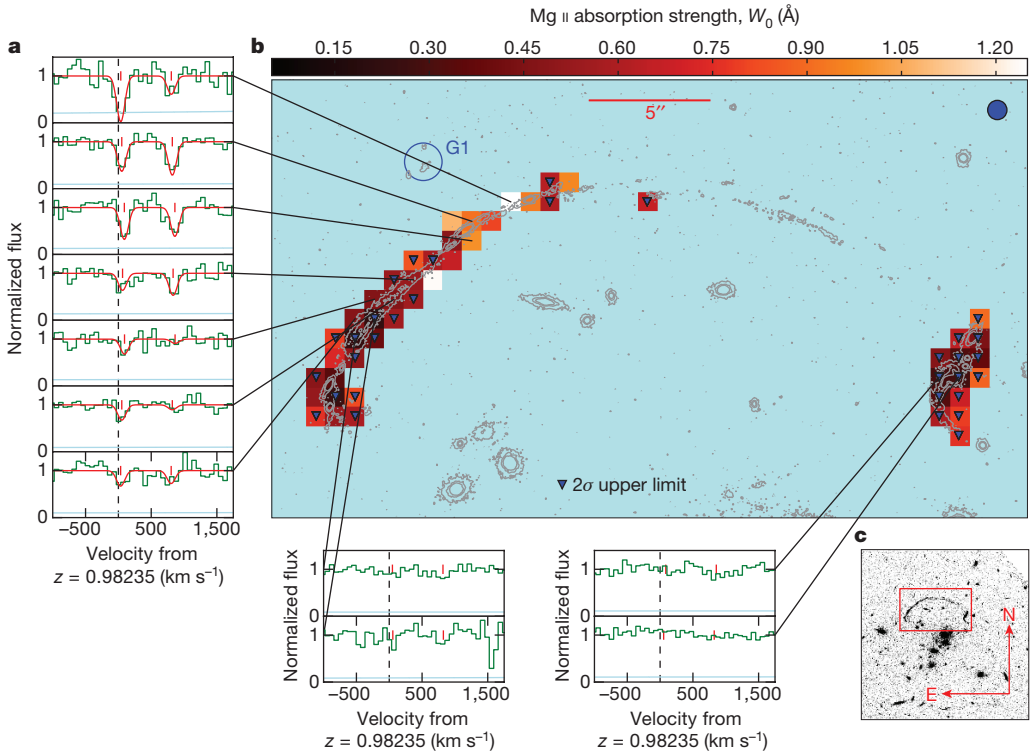


Figure 2 | Map of Mg II absorption strengths at approximately 4-kpc resolution. a, Representative sample of MUSE spectra of the Mg II $\lambda = 2{,}796$ Å, $\lambda = 2{,}803$ Å absorption doublets (green histograms) and corresponding Gaussian fits and line centroids (red lines and red ticks, respectively) in velocity space with respect to $z_{G1} = 0.98235$. Panels without fits correspond to non-detections. The black lines indicate the corresponding positions on the arc. **b**, $0.51' \times 0.27'$ inset of the $1' \times 1'$ MUSE field centred in the northeastern part of the arc. The colour scale indicates Mg II $\lambda = 2{,}796$ Å rest-frame absorption strengths obtained from the Gaussian fits. A total of 56 positions were selected to have a continuum signal-to-noise ratio above 3 at the expected Mg II lines (50 shown here), of which 18 have significant detections (see Extended Data Fig. 2, Extended Data Tables 1 and 2 and Methods). Non-detections are indicated by blue downward arrows. The candidate absorber G1 is indicated by the blue circle (to the northeast). For reference, we overlay arc and galaxy contours (grey lines) at 840 nm from HST data (GO programme 12267). Each independent spatial element (spaxel) is 0.8″ wide, equivalent to four MUSE unbinned spaxels and of similar size to the seeing (indicated by the blue circle in the top right). This grid is defined in the image plane; the actual spatial resolution varies across the absorber plane from about 4 kpc at the eastern side of the arc to about 2 kpc at the western side. Likewise, the 5″ scalebar corresponds to a range of roughly 24–12 kpc in the absorber plane, depending on position (see Extended Data Fig. 3 for a de-lensed image). **c**, Entire MUSE field of view indicating the location of the Mg II map shown in **b**.

In Fig. 2 we show a MUSE map of the $z_{abs} = 0.98$ Mg II absorption strength at different positions along RCSGA 032727−132623. In the absorber plane, the arc positions span projected physical distances ('impact parameters'; see Methods) of approximately 15–90 kpc to the absorbing galaxy system ('G1'; indicated by the blue circle). G1 is our primary absorber candidate of the three [O II] detections at redshift 0.98 (G1, G2 and G3; Methods) and deserves special attention. Hubble Space Telescope (HST) continuum images resolve this system into three irregular galaxies, G1-A, G1-B and G1-C, which all have blue $B - I$ colours. From the HST photometry we estimate that these galaxies have low luminosities (less than about 5% of L^*, the characteristic luminosity of galaxies) and consequently low stellar masses compared to quasar absorbers[3]. Such stellar masses suggest total halo masses of around 10^{11} solar masses ($M_\odot$), which in turn imply virial radii of the dark-matter haloes of approximately 90 kpc (Methods). In addition, from the [O II] $\lambda = 3{,}726$ Å, $\lambda = 3{,}729$ Å emission doublets, where λ is the wavelength, we estimate the star-formation rates of the G1 galaxies to be less than about $0.2 M_\odot$ yr^{-1}. We use the [O II] velocities to define a 'systemic' redshift for G1 of $z_{G1} = 0.98235$.

Four key features are readily evident from Fig. 2: (a) Mg II is detected from about 15 kpc out to about 45 kpc to the south of G1, but is not detected further than approximately 80 kpc to the west (another arc knot in the south, not shown in Fig. 2, also has sensitive non-detections; see below); (b) the absorption strength is not uniform, indicating a clumpy medium on 4-kpc scales down to our detection limit of approximately 0.4 Å; (c) the line centroids vary little (within one spectral pixel, or about 50 km s^{-1}) all across the arc; and (d) most of the doublet ratios appear to be saturated, indicating possible partial covering of the background source (Methods).

In Fig. 3 we show that the absorption strength decreases with impact parameter, in broad agreement with the quasar statistics[3], but that the scatter is smaller than in the quasar data[2,3]. The latter is probably a consequence of partial covering, which would skew down the arc measurements (Methods); however, the heterogeneity of the compiled quasar–galaxy sample may also have a role[3,19], because the intervening galaxies encompass a wide range of masses, luminosities and orientations. That is to say, we compare averages over different areas in a single galaxy (probed by the arc) with an average over many distinct galaxies (probed by quasars).

From quasar lens statistics[8,20] we know that transverse structure is detected on similar scales as probed by our 4-kpc, seeing-limited resolution, and below. This indicates that the metals traced by Mg II are concentrated in small clouds that we do not resolve here, but are spatially distributed in such a way as to produce the gradient that we observe. Therefore, our data do not probe individual cloud sizes but rather their coherence length. Some stringent non-detections (Fig. 3) re-enforce the notion of clumpiness on kiloparsec scales.

The tomographic technique that we use enables us to scan the velocity field of the absorbing gas profusely in a single high-redshift halo[10]. Figure 4 displays absorption velocities and emission velocities of G1-A, G1-B and G1-C as a function of impact parameter. The first outstanding feature in this figure is that all absorption velocities lie

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

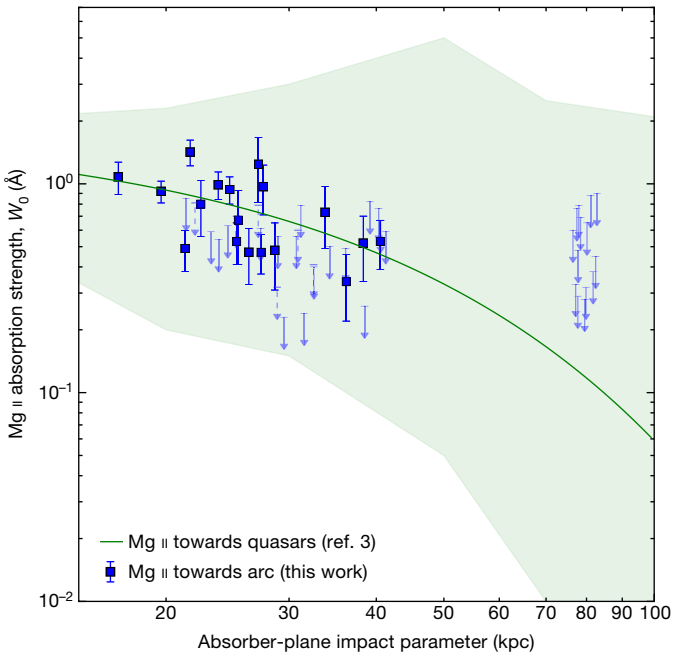


Figure 3 | Mg II absorption strength versus impact parameter. Squares correspond to detections and arrows to 2σ upper limits for the arc. Error bars correspond to the $\pm1\sigma$ uncertainty in the absorption strength measurement. All impact parameters are measured to G1 (northeast part of the arc), except for the six upper limits marked with dashed lines, which correspond to G2 (south). The zero-points are given by the centres of the blue circles in Fig. 2 and Extended Data Fig. 2. The green line is the fit to a sample of 182 quasar-absorption systems[3] and the shaded area indicates the full scatter in that sample. Uncertainties in the impact parameters are less than about 5%. 'Mg II towards quasars' and 'Mg II towards arc' in the legend indicates data obtained using quasars or the arc as background sources, respectively.

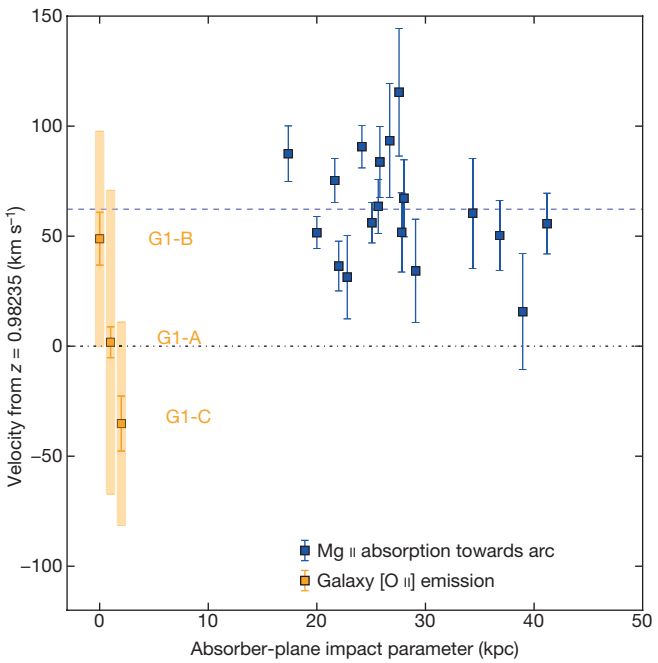


Figure 4 | Gas kinematics. Blue symbols correspond to the Mg II $\lambda = 2{,}796$ Å, $\lambda = 2{,}803$ Å absorption-line velocities of the arc and orange symbols to the [O II] $\lambda = 3{,}726$ Å and $\lambda = 3{,}729$ Å emission-line velocities of G1 (shifted in the x axis for clarity). The blue dashed line indicates the average absorption velocity. The zero-point velocity corresponds to the systemic redshift $z_{G1} = 0.98235$. Error bars correspond to the $\pm1\sigma$ uncertainty in the velocities. The envelopes around the emission-line measurements (orange shading) indicate the full-width at half-maximum (FWHM) velocity dispersion derived from the [O II] fits. Uncertainties in the impact parameters are less than about 5%.

redward of z_{G1}, at $+62$ km s^{-1}, although none of them substantially exceeds any of the velocity dispersions of the galaxies. Second, there is little overall variation in absorption velocity along the arc (approximately 24 km s^{-1}), even less than within the galaxy system itself, but this variation extends out to 40 kpc, a distance 10 times larger than the projected separations between the G1 galaxies. Along with a clumpy medium revealed by the map of absorption strengths, this kinematically quiet behaviour places strong constraints on the geometry and dynamics of this system.

Assuming saturation, the absorption strengths are a measure of the velocity spread of individual clouds[21] (not resolved by our data) and possible partial covering (see Methods). Therefore, most of our detections would correspond to velocity spreads of less than about 108 km s^{-1} along the sightlines, the equivalent of a 1-Å absorption line. Interestingly, we find lower scatter in the transverse direction. Taken together, these findings would indicate gas clouds with internal velocity dispersions that dominate over bulk motions. For the impact parameters and halo mass probed here, these velocities resemble those determined at higher spectral resolution in low-redshift systems[22,23], which appear well within the escape velocity and virial radius of the halo.

Our observations do not support a spherical geometry[24,25]. The Mg II gas does not seem to be distributed isotropically around G1; if this were the case, similar absorption would occur at both sides of the line connecting G1 and its closest arc position, which we do not see. Instead, there are more non-detections on the southern side. G2 (Methods) is also a good example of this situation: no Mg II is detected in six positions at roughly 20–30 kpc to sensitive limits, whereas four are expected if G2 followed the trend shown in Fig. 3 isotropically. Furthermore, assuming the observed Mg II to be related to only one of the G1 galaxies, then only one of the four galaxies (including G2) presents detectable absorption, which leads to a rough covering fraction of about 25% within 40 kpc. This interpretation assumes that the G1 galaxies are distinct objects and not part of a large disk of gas and dust (Methods). Therefore, in line with quasar-absorber observations[26,27], our arc observations strongly suggest that the absorbing gas is anisotropic, with wide (possibly around 90°) opening angles[28].

Our data also enable us to compute gas covering fractions in an alternative fashion, namely directly from our 'hits and misses' statistics around G1 only (Methods). Assuming anisotropy, we probe here the preferential G1 direction that shows absorption; therefore, our covering fraction estimate should lie above that of quasar absorbers, because those surveys probe random quasar–galaxy orientations. Interestingly, our prediction is fulfilled in comparison with galaxy-selected quasar absorbers[2], which we regard to be unbiased: the covering fraction towards the arc is indeed larger than towards the quasars in that sample (Methods). Conversely, a comparison with more heterogeneous samples that include Mg II-selected galaxies[3] yields smaller covering fractions towards the arc than towards the quasars. This suggests possible selection biases in the latter samples, because by construction they favour galaxy orientations that produce absorption[19].

At low redshift there is observational evidence[23,29] that the circumgalactic medium of star-forming galaxies is driven by the interplay between major-axis entraining gas (pristine or recycled) and enriched outflows aligned with the minor axis. Although this picture is also consistent with most recent simulations and ΛCDM predictions[30], it remains to be confirmed, particularly at high redshift. Here we deal with dwarf galaxies at redshift 1, which are still able to eject metals[12] out to one virial radius, beyond which the metal flux is expected to have decreased markedly[30]. The Mg II detections that we report occur well within the virial radius and therefore could have originated from any of the G1 galaxies (but given the relatively quiet velocity field, most probably from only one of them). In addition, the gas is metal-enriched

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

(including in Fe II; see Methods) and patchy (perhaps revealing a multi-phase medium), which is suggestive of recycled gas either outflowing from or bound to one of the individual G1 haloes. The outflow scenario requires outflow speeds[6,26] in excess of the offsets between the velocities of the galaxies and the gas observed here (Fig. 4). This suggests that the Mg II is more likely to be correlated with entrained gas in the extended halo of G1-B (closer in velocity), or even G1-A, in which case the absorption velocity offset would resemble the one-sided velocity offsets seen at low redshift[22,23].

The gravitational-arc tomography presented here appears to probe the gaseous extension of a galaxy halo in formation, at distances greater than 20 kpc, which might be a remnant of past gravitational interactions that formed tidal debris and gaseous streams infalling back into the overall G1 potential well. Unfortunately, the arc-galaxy configuration under study does not cover lower impact parameters for testing this hypothesis, but future objects may permit more conclusive detections.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 July; accepted 14 November 2017.
Published online 31 January 2018.

1. Churchill, C. W. *et al.* Low- and high-ionization absorption properties of Mg II absorption-selected galaxies at intermediate redshifts. I. General properties. *Astrophys. J. Suppl. Ser.* **130,** 91–119 (2000).
2. Chen, H.-W. *et al.* An empirical characterization of extended cool gas around galaxies using Mg II absorption features. *Astrophys. J.* **714,** 1521–1541 (2010).
3. Nielsen, N. M., Churchill, C. W. & Kacprzak, G. G. MAGIICAT II. General characteristics of the Mg II absorbing circumgalactic medium. *Astrophys. J.* **776,** 115 (2013).
4. Chen, H.-W. in *Outskirts of Galaxies* (eds Knapen, J. *et al.*) 291–331 (Springer, 2017).
5. Tytler, D. *et al.* Metal absorption systems in spectra of pairs of QSOs: how absorbers cluster around QSOs and other absorbers. *Mon. Not. R. Astron. Soc.* **392,** 1539–1572 (2009).
6. Martin, C. L. *et al.* The size and origin of metal-enriched regions in the intergalactic medium from spectra of binary quasars. *Astrophys. J.* **721,** 174–192 (2010).
7. Hennawi, J. F. *et al.* Binary quasars in the Sloan Digital Sky Survey: evidence for excess clustering on small scales. *Astron. J.* **131,** 1–23 (2006).
8. Rauch, M., Sargent, W. L. W., Barlow, T. A. & Carswell, R. F. Small-scale structure at high redshift. III. The clumpiness of the intergalactic medium on subkiloparsec scales. *Astrophys. J.* **562,** 76–87 (2001).
9. Lopez, S., Ellison, S., D'Odorico, S. & Kim, T.-S. Clues to the nature of high-redshift O VI absorption systems from their lack of small-scale structure. *Astron. Astrophys.* **469,** 61–74 (2007).
10. Chen, H.-W. *et al.* Spatially resolved velocity maps of halo gas around two intermediate-redshift galaxies. *Mon. Not. R. Astron. Soc.* **438,** 1435–1450 (2014).
11. Rubin, K. H. R. *et al.* Andromeda's parachute: a bright quadruply lensed quasar at $z=2.377$. Preprint at https://arxiv.org/abs/1707.05873 (2017).
12. Kereš, D., Katz, N., Weinberg, D. H. & Davé, R. How do galaxies get their gas? *Mon. Not. R. Astron. Soc.* **363,** 2–28 (2005).
13. Muratov, A. L. *et al.* Gusty, gaseous flows of FIRE: galactic winds in cosmological simulations with explicit stellar feedback. *Mon. Not. R. Astron. Soc.* **454,** 2691–2713 (2015).
14. Bacon, R. *et al.* The MUSE second-generation VLT instrument. *Proc. SPIE* **7735,** 773508 (2010).
15. Wuyts, E. *et al.* A bright, spatially extended lensed galaxy at $z=1.7$ behind the cluster RCS2 032727–132623. *Astrophys. J.* **724,** 1182–1192 (2010).
16. Dahle, H. *et al.* Discovery of an exceptionally bright giant arc at $z=2.369$, gravitationally lensed by the Planck cluster PSZ1 G311.65–18.48. *Astron. Astrophys.* **590,** L4 (2016).
17. Rigby, J. R., Wuyts, E., Gladders, M. D., Sharon, K. & Becker, G. D. The physical conditions of a lensed star-forming galaxy at $z=1.7$. *Astrophys. J.* **732,** 59 (2011).
18. Rigby, J. R. *et al.* The Magellan Evolution of Galaxies Spectroscopic and Ultraviolet Reference Atlas (MEGaSaURA) I: the sample and the spectra. *Astrophys. J. Suppl. Ser.* (in the press).
19. Nielsen, N. M., Churchill, C. W., Kacprzak, G. G. & Murphy, M. T. MAGIICAT I. The Mg II absorber-galaxy catalog. *Astrophys. J.* **776,** 114 (2013).
20. Lopez, S. *et al.* Metal abundances in a damped Lyα system along two lines of sight at $z=0.93$. *Astrophys. J.* **626,** 767–775 (2005).
21. Ellison, S. L. An efficient technique for pre-selecting low-redshift damped Lyα systems. *Mon. Not. R. Astron. Soc.* **368,** 335–340 (2006).
22. Tumlinson, J. *et al.* The COS-Halos Survey: rationale, design, and a census of circumgalactic neutral hydrogen. *Astrophys. J.* **777,** 59 (2013).
23. Ho, S. H., Martin, C. L., Kacprzak, G. G. & Churchill, C. W. Quasars probing galaxies. I. Signatures of gas accretion at redshift approximately 0.2. *Astrophys. J.* **835,** 267 (2017).
24. Steidel, C. C. in *QSO Absorption Lines (*ed. Meylan, G.) 139–152 (Springer, 1995).
25. Charlton, J. C. & Churchill, C. W. Mg II absorbing galaxies: halos or disks? *Astrophys. J.* **465,** 631–645 (1996).
26. Bouché, N. *et al.* Physical properties of galactic winds using background quasars. *Mon. Not. R. Astron. Soc.* **426,** 801–815 (2012).
27. Kacprzak, G. G., Churchill, C. W. & Nielsen, N. M. Tracing outflows and accretion: a bimodal azimuthal dependence of Mg II absorption. *Astrophys. J.* **760,** L7 (2012).
28. Bordoloi, R., Lilly, S. J., Kacprzak, G. G. & Churchill, C. W. Modeling the distribution of Mg II absorbers around galaxies using background galaxies and quasars. *Astrophys. J.* **784,** 108 (2014).
29. Kacprzak, G. G. in *Gas Accretion onto Galaxies* (eds Fox, A. & Davé, R.) 145–165 (Springer, 2017).
30. Muratov, A. L. *et al.* Metal flows of the circumgalactic medium, and the metal budget in galactic haloes. *Mon. Not. R. Astron. Soc.* **468,** 4170–4188 (2017).

Acknowledgements This work has benefitted from discussions with A. Smette, N. Nielsen and G. Kacprzak. S.L. thanks the European Southern Observatory Scientific Visitor Selection Committee for supporting a research stay at the ESO headquarters in Santiago, where part of this work was done. S.L. has been supported by FONDECYT grant number 1140838. This work has also been partially supported by PFB-06 CATA. N.T. acknowledges support from CONICYT PAI/82140055.

Author Contributions S.L. conceived and led the project. S.L. and N.T. wrote the MUSE telescope-time proposal and designed the observations. L.F.B. and N.T. prepared the remote observations and L.F.B. reduced the MUSE data. S.L., N.T. and C.L. analysed the data, performed simulations and devised ways to produce and interpret the results. S.L. wrote the main codes. N.T. and I.P. performed the blind survey of galaxies in the field of view. K.S. performed the lens model and L.F.B. supervised the design of Fig. 1. M.B.B. and L.F.B. performed the photometric characterization of the absorbing galaxies, and S.L., C.L. and N.T. the analysis of their spectra. Ancillary data from MagE and HST were provided by J.R.R. and M.D.G. S.L. wrote the manuscript and produced the rest of the figures, with contribution from N.T. All co-authors provided critical feedback and helped to shape the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to S.L. (slopez@das.uchile.cl).

Reviewer Information *Nature* thanks H.-W. Chen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Observations and data reduction. The data were obtained with the MUSE integral-field spectrograph[14] mounted on the ESO Very Large Telescope. The 1′ field of view is sampled with 349 × 352 0.2″-wide spaxels. MUSE uses image slicers combined with an array of 24 identical spectrographs that provide nearly 100,000 spectra simultaneously. Our set-up provided a wavelength range of 4,650–9,300 Å at a resolving power $R=\lambda/\Delta\lambda$ of 2,000–4,000. Each spectral bin is 1.25-Å wide. The observations were carried out in 'service mode' during dark time, with thin cirrus or clear-sky conditions, airmass below 1.8 and seeing better than 0.8″. We obtained a total of 21 exposures of 700-s on-target each. The exposures were taken within 'observing blocks' of four exposures each. We applied a small dithering and 90° rotations between exposures to reject cosmic rays and minimize patterns of the slicers on the processed combined cube. We reduced all of the observations using the MUSE pipeline recipe v1.6.4 and ESO reflex v2.8.5. We analysed the quality of the individual exposures by measuring the seeing and transparency in the 'white' images. Owing to changes in the weather conditions during integrations, some of the exposures come from aborted observing blocks (but are otherwise 700-s long). We discarded five of them for having guiding errors, poor point spread function or obvious transparency problems. The remaining 16 exposures were combined into one final science datacube. The total on-target time was therefore 3.1 h. The sky subtraction was improved on this cube using the Zurich Atmospheric Purge (ZAP) algorithm[31]. The wavelength solution, corrected for the motion of the Earth around the Sun and converted to vacuum, was checked satisfactorily at the position of the brightest knot in the arc against our MagE spectrum[18]. The spectral resolution at the Mg II absorption lines is FWHM = 2.7 Å, corresponding to about 140 km s^{-1} (or $R=2{,}100$).

In this work we adopt a flat ΛCDM cosmology with $\Omega_\Lambda=0.7$, $\Omega_m=0.3$ and $H_0=70$ km s^{-1} Mpc^{-1}.

Spaxel combination. Unlike quasars, the source we study is (a) extended and (b) inhomogeneous in flux given the resolved background galaxy and the inhomogeneous lensing magnifications (Extended Data Fig. 1). Therefore, the spectrum extraction must proceed using an ad hoc flux combination to assure independent flux measurements while maximizing the final signal-to-noise ratio per spectral pixel. To first approximation we tried polygons of constant signal-to-noise ratio[32], but decided not to use them because widely different areas complicate the interpretation of absorption against an extended source. Instead, we perform a 'minimum size' binning based on a fixed grid that was defined by the spatial sampling. Because the seeing profile is sampled with four 0.2″ spaxels (FWHM) we use 4 × 4 binning. This means that the centre of each binned spaxel lies at least one seeing FWHM from any other. Offsetting the grid by ±1 unbinned spaxel shows that the choice of zero-point does not affect the results of the analysis.

Spectral extraction. To extract individual one-dimensional spectra for each binned spaxel (or 'position') we applied weights $w_{ij}=f_{ij}/v_{ij}$, where f_{ij} and v_{ij} are respectively the calibrated object flux and variance of spaxel (i, j). This weight is a modified version of that used to combine spectra with different signal-to-noise ratios optimally[33]; it assumes a low detector read-out noise, in which case v_{ij} approximates the total number of counts. This gives higher weights to brighter spaxels and may introduce a complicated bias towards highly magnified spaxels. Because we are interested in a small spectral range we defined a subcube that contains wavelengths 5,100–6,000 Å. This range includes the Mg II doublet, the Fe II $\lambda=2{,}382$ Å and Fe II $\lambda=2{,}600$ Å transitions, and the Mg I $\lambda=2{,}852$ Å transition at $z_{\rm abs}=0.98$. Because the array of variances is noisy in the spectral direction, to avoid introducing noise carried by the spectral weights we choose a weight integrated in the spectral direction in a small featureless region blueward of the Mg II lines (the red side is compromised by a sky emission-line residual). Also, this weight does not penalize the absorption troughs.

Spaxel selection and Gaussian fits. To search for Mg II we first created a signal-to-noise ratio map by selecting binned spaxels on top of the arc and with signal-to-noise ratio above 3 blueward of the expected Mg II absorption. This selected a total of 56 positions, corresponding to a total projected area surveyed of around 600 kpc^2. At each selected position a spectrum was extracted and an automatic Gaussian fit applied to the spectral region corresponding to $z_{\rm G1}=0.98235$. The Mg II $\lambda=2{,}796$ Å, $\lambda=2803$ Å doublet was fitted with two Gaussians with a tied wavelength ratio, free doublet ratio and fixed line width (corresponding to the instrument spectral resolution, FWHM = 2.7 Å). Each fit provides a rest-frame absorption strength of the $\lambda=2{,}796$ Å transition (or rest-frame equivalent width 'W_0') and a redshift (or radial velocity v), along with their statistical errors (Extended Data Tables 1 and 2). Fixing the line width provides more robust fits, as expected, avoiding false positives in regions with low signal-to-noise ratio. It also assumes that the instrumental profile dominates the line profile. This assumption may not hold for all of the positions, in which case we estimate that a systematic error of approximately 0.05 Å would be introduced in W_0 (not included in Extended Data Tables 1 and 2).

In addition to the fits, synthetic line profiles were created for comparison with the data. When a fit failed or the significance was below 3σ, a 2σ upper limit was calculated using $\sigma_{W_0}(1+z)=\mathrm{FWHM}/\langle S/N\rangle$, where σ_{W_0} is the expected rest-frame 1σ error in the W_0 measurement and $\langle$S/N$\rangle$ is the average continuum signal-to-noise ratio near the line. The procedure delivers 18 significant Mg II detections, all to the northeast of the arc. Finally, to create the absorption-strength and velocity maps, the fitted W_0 and v values (or the W_0 upper limits) were recorded in images that had the same spatial sampling as the signal-to-noise ratio map.

We attempted simultaneous fits to the Fe II $\lambda=2{,}382$ Å and Mg I $\lambda=2{,}852$ Å transitions tied to the Mg II redshifts. (Unfortunately, the Fe II $\lambda=2{,}600$ Å transition is blended with the C III] $\lambda=1{,}907$ Å, $\lambda=1{,}909$ Å emission-line doublet from the source galaxy[34] and therefore could not be fitted.) We detect significant Fe II at only three positions along the arc (Extended Data Table 3). These positions also show the strongest and most significant Mg II absorption. The corresponding Fe II/Mg II strength ratios lie in the rather narrow range of 0.5–0.7 and conform to the quasar statistics of very strong ($W_0>1$ Å) Mg II systems[35]. Non-detections do not constrain these ratios at other positions of the arc owing to either too weak Mg II or too low signal-to-noise ratio. Although the Fe II data are limited, the similar ratios tentatively hint at homogeneous enrichment along the arc. One of these positions shows a (marginal) Mg I detection.

The absorption signal in the maps, although significant, is expected to be weak and directly affected by the inhomogeneous signal-to-noise ratio. To rule out possible artefacts due to reduction or analysis effects, we conducted mock tests by simulating a flat W_0 distribution in a cube with the same signal-to-noise ratio per spaxel as the actual data. The outcome of these tests shows that our fitting procedure recovers a true signal in nearly 100% of the (binned) spaxels with a signal-to-noise ratio above 3, which justifies our signal-to-noise ratio cutoff.

Galaxies at $z=0.98$. We searched systematically for galaxies in the MUSE cube near $z=0.98$. The search included continuum sources and emission-line galaxies. We detected three [O II] sources, which we refer to as G1, G2 and G3 (Extended Data Fig. 2). These sources form a triangle with sides of 42″, 47″ and 64″, or $d=231$ kpc, 259 kpc and 341 kpc in the absorber plane. G1 is resolved into three galaxies in the HST continuum images, which are barely resolved by MUSE. We refer to them as G1-A, G1-B and G1-C. From Gaussian fits to the [O II] $\lambda=3{,}726$ Å, $\lambda=3{,}729$ Å doublet in the MUSE spectra we obtained redshifts, line velocity dispersions, [O II] luminosities and star-formation rates for the five galaxies. The deconvolved velocity dispersions are subject to systematics, owing to the modest MUSE resolution. Assuming virial equilibrium, the velocities of G1, G2 and G3 lead to a virial radius $R_{\rm vir}<d$ and thus we do not consider them to be bound gravitationally, owing to their projected separations. Instead, we deem them to be three independent systems that lie in the same large-scale structure at $z=0.98$.

G1-A, G1-B and G1-C represent our best absorbing-galaxy-system candidate, owing to their proximity to the place where the arc absorption occurs. The proximity of G1-A, G1-B and G1-C may cast doubts on whether they are indeed distinct galaxies, as opposed to a single disk in which dust obscuration would mimic the presence of different objects. But the fact that they are resolved also in the rest-frame I band (F160W) supports the existence of distinct galaxies.

We obtained the photometry from HST images in the F606W, F814W, F098M, F125W and F160W bands. We used SExtractor[36] in dual mode using the detections in the F160W band as a reference to obtain AB magnitudes in a 0.8″-diameter aperture. $B-I$ rest-frame colours were computed from F814W and F160W because these filters are the closest in effective wavelength. We used a local Scd galaxy spectral template[37] that well represents the colour of these galaxies to correct for any mismatch in the effective passbands. Small extinction corrections[38] were also applied. Using the multi-band photometry and a spectral energy distribution (SED) fitting code[39], we estimated luminosities, stellar masses and star-formation rates. These quantities are subject to large uncertainties owing to the use of just five passbands. We also computed star-formation rates for each galaxy from the [O II] line fluxes, integrated over 16 unbinned spaxels. These line fluxes are subject to extinction and therefore must be treated only as lower limits. Using Kennicutt's prescription[40], the emission-line luminosities translate into star-formation rates that are broadly consistent with those obtained from the SED fitting. These values were corrected for magnification μ by using our lens model. From the stellar masses we estimate dark-matter halo masses $M_{\rm h}$ using the prediction from abundance matching[41]. We then determine the corresponding virial radius using the relation $R_{\rm vir}=[3M_{\rm h}/(200\rho_{\rm c}4\pi)]^{1/3}$, where $\rho_{\rm c}$ is the critical density of the Universe at $z=0.98$. The galaxy data are summarized in Extended Data Table 4.

Lens model. We base our lens model on a previous lensing analysis of this cluster[42]. We improve this previous lens model with new lensing constraints, including three new spectroscopic redshifts that we measured from the MUSE data. We obtain a spectroscopic redshift for the radial arc S7a/S7b in ref. 42 of $z_{\rm spec}=2.82624$. We identify two new sets of lensed galaxies, at $z_{\rm spec}=2.7$ and $z_{\rm spec}=5.2$. Including more spectroscopic redshift constraints substantially reduces the lens model

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

uncertainties and improves the accuracy[42,43]. The lens model is computed using the public software Lenstool[44], which uses a Markov chain Monte Carlo (MCMC) process to explore the parameter space. The lens model results in a computed projected mass density distribution in the lens plane, magnification maps for any given redshift, deflection fields, and their uncertainties. The deflection matrices are calculated using the lens equation, $\boldsymbol{\beta} = \boldsymbol{\theta} - (d_{ls}/d_s)\boldsymbol{\alpha}(\boldsymbol{\theta})$, where $\boldsymbol{\beta}$ is the source position at z_{source}, $\boldsymbol{\theta}$ is the observed position, d_{ls} and d_s are the distances from the lens to the source and from the observer to the source, respectively, and $\boldsymbol{\alpha}(\boldsymbol{\theta})$ is the deflection angle at the observed position. We note that any plane behind the lens can be considered the source plane, and this equation also applies to the absorber plane.

To assess the completeness of our search for [O II] in emission, we scanned the magnification map near Mg II looking for regions with much lower magnification than on top of G1 and G2. We found none, indicating that these galaxies are not sitting on particularly highly magnified regions; consequently, we are confident that our absorber candidates are robust and no other galaxies (of similar brightness but non-magnified) were missed from our search.

Spatial resolution and impact parameters. We use the deflection matrices to de-lens the coordinates of our binned spaxels into the absorber plane and to calculate impact parameters. Extended Data Fig. 3 shows a de-lensed image of the arc projected against the image plane. Owing to the inhomogeneous lensing deflection, when the spaxels are traced from the image plane to the absorber plane the shape of the binned spaxels changes in the absorber plane, although there is no overlap between them. Assuming that the light rays do not intercept each other after being absorbed, our signal should probe independent areas of the absorber. We discuss the unequal spaxel areas below.

From Extended Data Fig. 3 it is also evident that although the angular resolution is constant across the image plane, the actual spatial resolution—defined in the absorber plane—is not. To define an ad hoc 'spatial resolution' we take the square root of the area of each de-lensed spaxel. We find that this number ranges from around 4 kpc at the east side of the arc to around 2 kpc at the western side.

To calculate impact parameters we multiply the de-lensed angular separations between spaxels and G1 by the scale at $z = 0.98$ given by the adopted cosmology, 7.97 kpc per arcsecond. From a large set of MCMC realizations, we estimate the statistical 1σ uncertainty associated with the angular separations to be less than about 2%. Including model systematics[43], impact parameters should be precise to less than about 5% for the assumed cosmology.

Partial covering. The background source is likely to be more extended than the typical size of the absorbing clouds, leading to possible partial covering[45]. To test this effect on our signal we performed a second run of automatic Mg II fits on a 8×8 spaxel map. Extended Data Fig. 4 shows the cumulative distributions of W_0 and v for the 4×4 and 8×8 binnings. The stronger binning indeed skews the W_0 distribution to lower values. This is probably due to averaging arc light without Mg II absorption; although, given the arc geometry, at this spaxel size (1.6″) we also expect some sky contamination (not so in the 4×4 binning). Conversely, the velocity distribution remains unaffected. We conclude that (a) the chosen 4×4 binning is our best option above the seeing limitation; and (b) we cannot exclude a level of partial covering in our W_0 sample, although such an effect is not affecting the absorption velocities.

Particularly relevant to interpreting our results, the physical areas covered by the binned spaxels are unequal in the absorber plane (Extended Data Fig. 3). These areas are of the order of 10 kpc^2, at least 10^7 times larger than those probed by the quasar beams (less than about 1 pc). Clearly, our arc measurements sample an average signal at each position. Therefore, measurements at different positions are comparable with each other and, to a great extent, also independent of the spaxel area. The only effect of having uneven areas in the absorber plane should be on the intrinsic scatter of each measurement, with smaller spaxels having more scatter. We cannot measure this intrinsic scatter, but consider its effect to be negligible given that the goal is to make a comparison with measurements obtained along the much narrower quasar beams.

Gas covering fractions. To calculate gas covering fractions we compute the fraction of positions with positive detections above a given W_0 cutoff, W_{cut}, in a given impact parameter (D) range. Non-detections are accounted for by considering only (2σ) upper limits below W_{cut}. We select W_{cut} and ranges of D to enable comparisons with two quasar-absorber statistics.

To compare with the survey presented in ref. 2, we assume an average galaxy magnitude $\langle M_B \rangle = -19.0$ and use $W_{cut} = 0.5$ Å. We find covering fractions of 100% (2/2; in the range 16 kpc $< D <$ 20 kpc), 80% (4/5; 20 kpc $< D <$ 25 kpc) and 38% (6/16; 25 kpc $< D <$ 39 kpc). The quasar statistics for low-luminosity galaxies gives[2] 60%, 20% and 0%, respectively, for the three ranges of D. These figures suggest that the covering fraction of the intervening gas is larger towards this arc than towards the quasars in that sample.

On the other hand, to compare with the survey presented in ref. 3, we use two cut-offs: $W_{cut} = 0.6$ Å and $W_{cut} = 1.0$ Å. For $W_{cut} = 0.6$ Å we find covering fractions of 67% (6/9; 0 kpc $< D <$ 25 kpc), and 20% (4/20; 25 kpc $< D <$ 50 kpc). The quasar statistics for low-luminosity galaxies ($0.07 < L_B/L_B^* < 0.94$, where L^* is the characteristic luminosity of galaxies and B denotes the rest-frame B band) gives[3] 71% and 48%, respectively. For $W_{cut} = 1.0$ Å we find covering fractions of 18% (2/11) and 4% (1/26), respectively, for the same ranges of D. The quasar statistics gives 29% and 24%. Therefore, for both W_0 cut-offs, the covering fractions for the arc are smaller than those for the quasars in this sample.

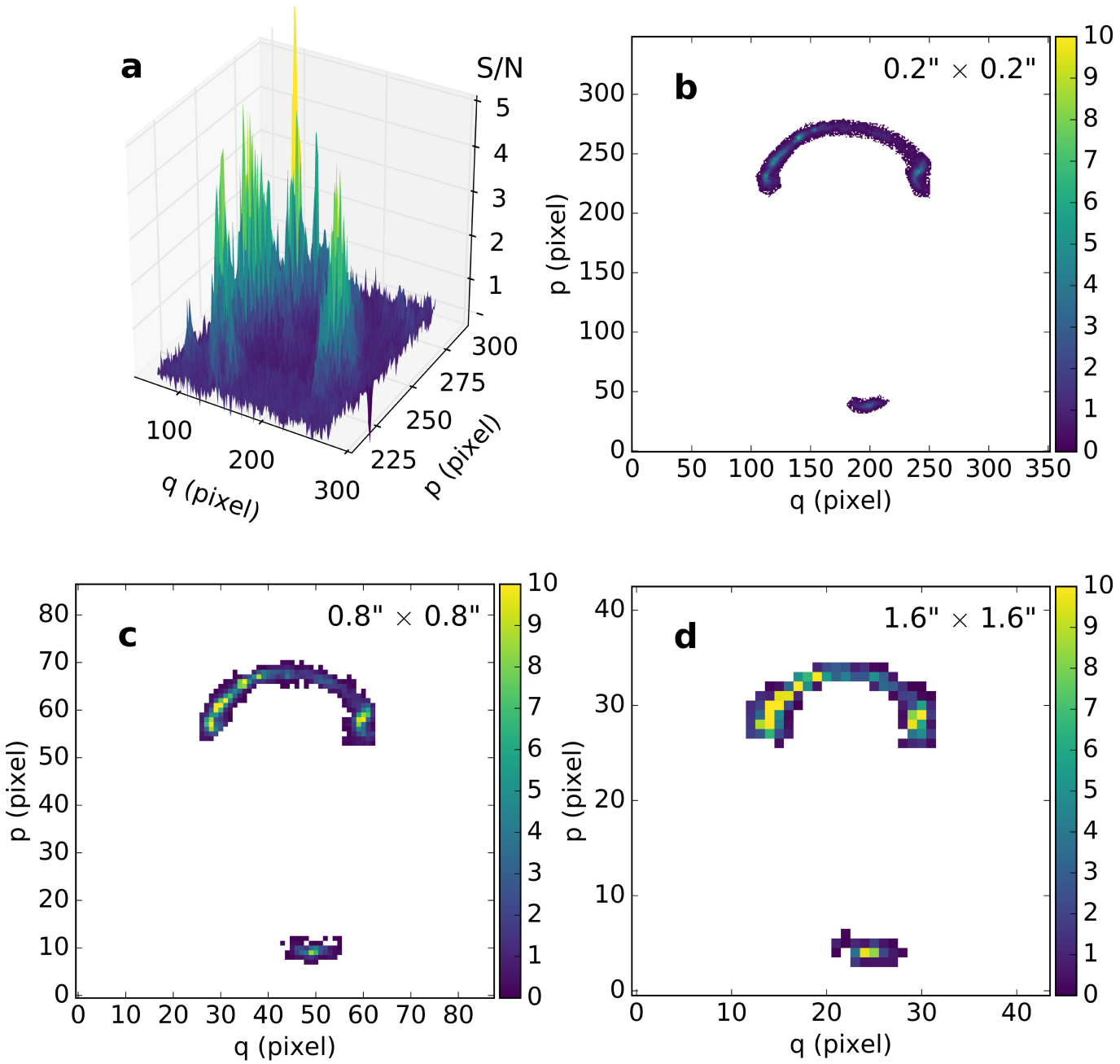
Finally, we note that these comparisons are subject to uncertainties because we do not cover exactly the same impact-parameter ranges.

Code availability. This analysis is based on custom Python routines, some of which use the MUSE Python data analysis framework[46] and the open-source plotting package for Python APLpy[47]. We have opted not to make our routines available because they are described in detail in the paper.

Data availability. The observations discussed in this paper were made using European Southern Observatory (ESO) Telescopes at the La Silla Paranal Observatory under programme ID 098A.0459(A). The corresponding data are available on the ESO archive at http://archive.eso.org/cms.html.

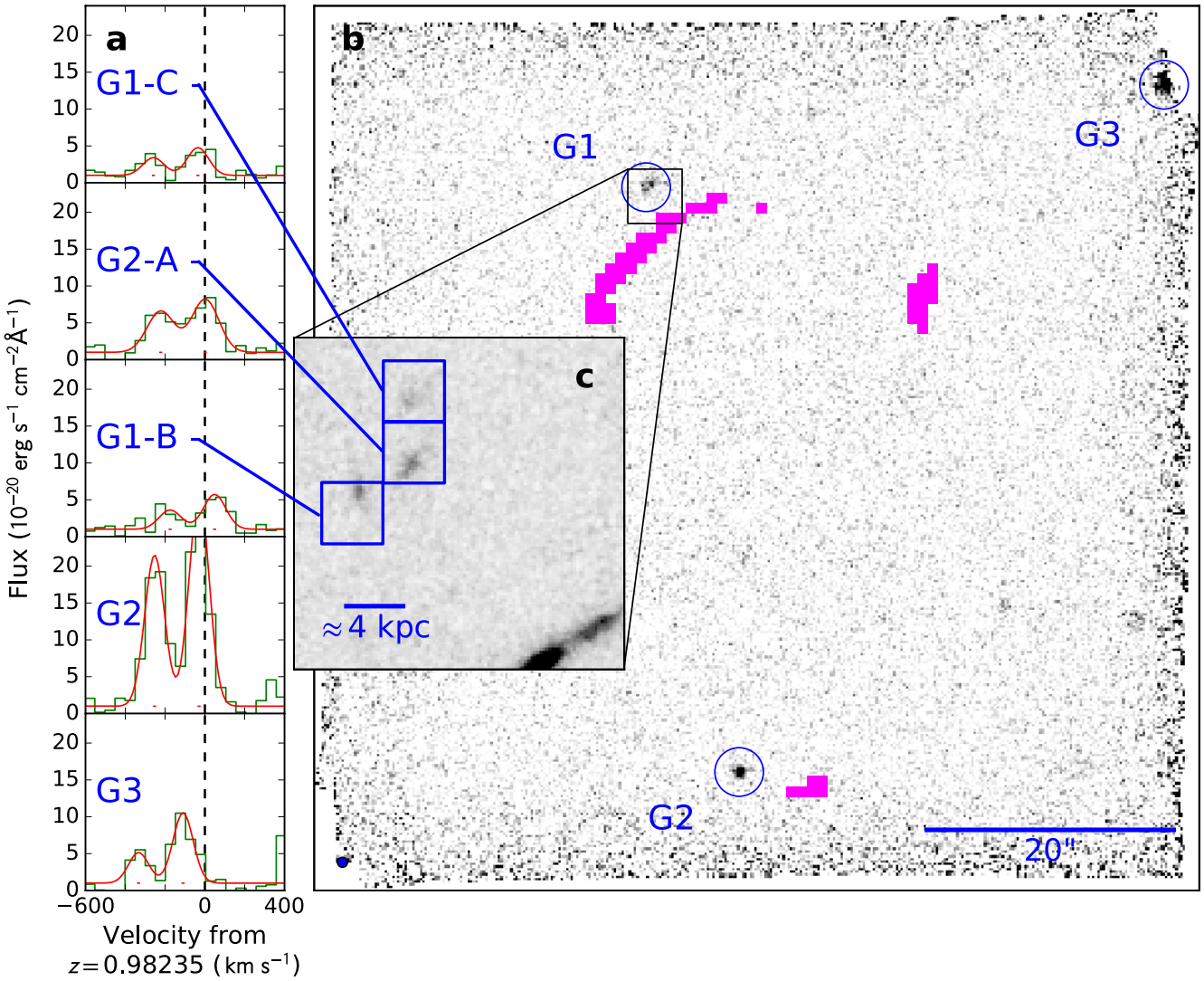
31. Soto, K. T., Lilly, S. J., Bacon, R., Richard, J. & Conseil, S. ZAP – enhanced PCA sky subtraction for integral field spectroscopy. *Mon. Not. R. Astron. Soc.* **458,** 3210–3220 (2016).
32. Cappellari, M. & Copin, Y. Adaptive spatial binning of integral-field spectroscopic data using Voronoi tessellations. *Mon. Not. R. Astron. Soc.* **342,** 345–354 (2003).
33. Robertson, J. G. Optimal extraction of single-object spectra from observations with two-dimensional detectors. *Publ. Astron. Soc. Pacif.* **98,** 1220–1231 (1986).
34. Rigby, J. R. *et al.* C III emission in star-forming galaxies near and far. *Astrophys. J. Lett.* **814,** L6 (2015).
35. Rodríguez Hidalgo, P. *et al.* Evolution of the population of very strong Mg II absorbers. *Mon. Not. R. Astron. Soc.* **427,** 1801–1815 (2012).
36. Bertin, E. & Arnouts, S. SExtractor: software for source extraction. *Astron. Astrophys. Suppl. Ser.* **117,** 393–404 (1996).
37. Coleman, G. D., Wu, C.-C. & Weedman, D. W. Colors and magnitudes predicted for high redshift galaxies. *Astrophys. J. Suppl. Ser.* **43,** 393–416 (1980).
38. Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500,** 525–553 (1998).
39. Moustakas, J. *et al.* PRIMUS: constraints on star formation quenching and galaxy merging, and the evolution of the stellar mass function from $z = 0$–1. *Astrophys. J.* **767,** 50 (2013).
40. Kennicutt, R. C. Jr. Star formation in galaxies along the Hubble sequence. *Annu. Rev. Astron. Astrophys.* **36,** 189–231 (1998).
41. Moster, B. P. *et al.* Constraints on the relationship between stellar mass and halo mass at low and high redshift. *Astrophys. J.* **710,** 903–923 (2010).
42. Sharon, K. *et al.* Source-plane reconstruction of the bright lensed galaxy RCSGA 032727–132609. *Astrophys. J.* **746,** 161 (2012).
43. Johnson, T. L. & Sharon, K. The systematics of strong lens modeling quantified: the effects of constraint selection and redshift information on magnification, mass, and multiple image predictability. *Astrophys. J.* **832,** 82 (2016).
44. Jullo, E. *et al.* A Bayesian approach to strong lensing modelling of galaxy clusters. *New J. Phys.* **9,** 447 (2007).
45. Bergeron, J. & Boissé, P. Extent and structure of intervening absorbers from absorption lines redshifted on quasar emission lines. *Astron. Astrophys.* **604,** A37 (2017).
46. Bacon, R., Piqueras, L., Conseil, S., Richard, J. & Shepherd, M. MPDAF: MUSE Python data analysis framework. *Astrophysics Source Code Library* (2016).
47. Robitaille, T. & Bressert, E. APLpy: astronomical plotting library in Python. *Astrophysics Source Code Library* (2012).
48. Willmer, C. N. A. *et al.* The Deep Evolutionary Exploratory Probe 2 galaxy redshift survey: the galaxy luminosity function to $z \sim 1$. *Astrophys. J.* **647,** 853–873 (2006).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



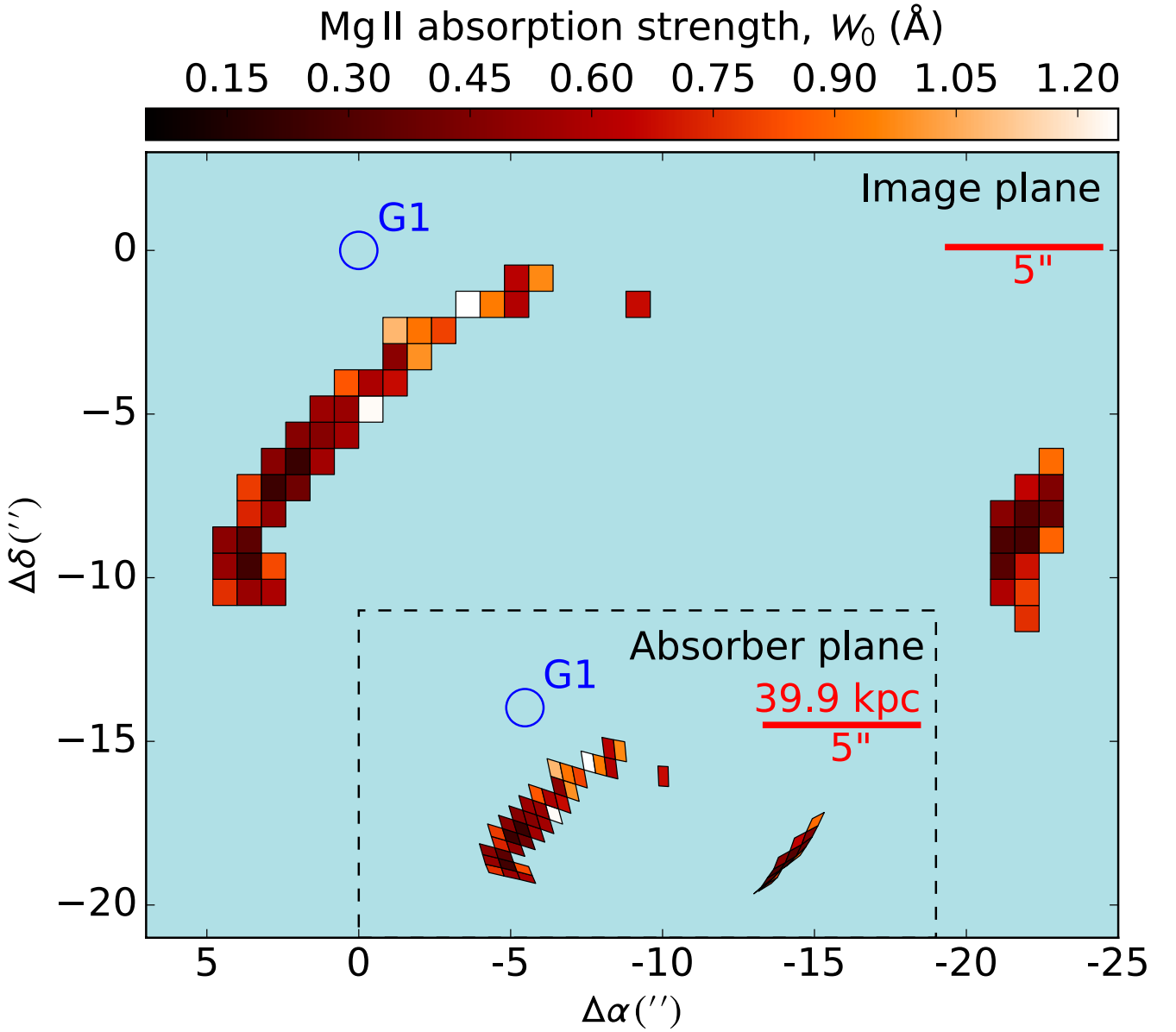
Extended Data Figure 1 | Signal-to-noise ratio versus binning. **a**, Three-dimensional representation of the signal-to-noise ratio (S/N) at the position of Mg II absorption in the unbinned data. **b**–**d**, Same as **a** but in two dimensions and for different binnings. The size of each binned spaxel is indicated in arcseconds; the colour scale is the same for all three panels. Note the expected increase in signal-to-noise ratio with binning size.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



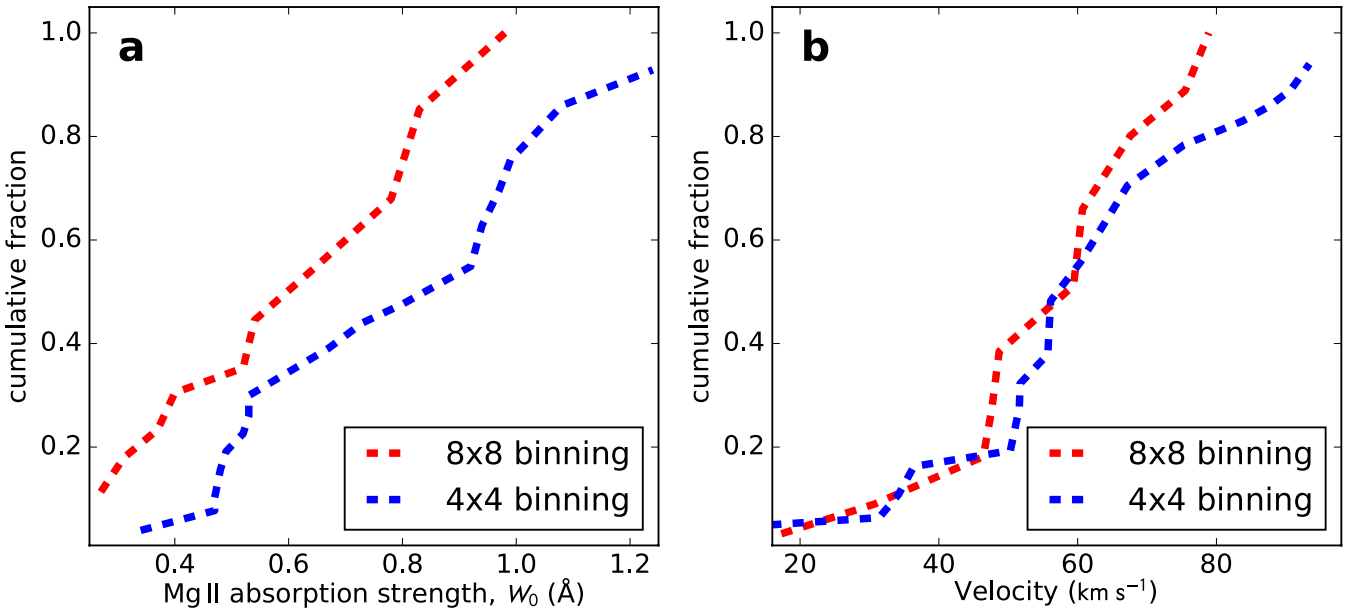
Extended Data Figure 2 | Emission-line galaxies at $z = 0.98$. a, Gaussian fits to the [O II] $\lambda = 3{,}726$ Å, $\lambda = 3{,}729$ Å doublets in the MUSE spectra of G1, G2 and G3, the three [O II] sources found by our systematic search. The MUSE spatial resolution barely resolves G1 into three [O II] clumps (G1-A, G1-B and G1-C), which cluster around $z_{G1} = 0.98235$ and have a velocity dispersion of 35 km s^{-1}. **b**, MUSE image of RCSGA 032727−132623 centred on [O II] emission at $z = 0.98$. The magenta squares indicate the binned spaxels used to map the Mg II absorption against the arc. **c**, HST/WFC3 F814W image zooming into the G1 system. The blue squares indicate the MUSE regions used to extract the [O II] spectra. The scale corresponds to the region close to G1 in the absorber plane.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 3 | Projection of absorber plane against image plane. In the absorber plane (dashed rectangle), the spaxel configuration appears shrunken and the de-lensed spatial elements have different shapes and areas across the absorber plane. After de-lensing, the scale in the absorber plane is given by the adopted cosmology: $5'' = 39.85$ kpc at $z = 0.98$. The impact parameter used here is defined as the projected physical distance between a given position and G1 on this plane. For reference, a $5''$ scale bar is shown in the image plane. Coordinates (α, right ascension; δ, declination) are in arcseconds relative to the position of G1 in the image plane.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 4 | Effect of partial covering. a, Cumulative distribution of absorption strengths for two different binnings. **b**, Same as **a** but for velocities.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Mg II absorption near G1

$\Delta\alpha^*$ ($''$)	$\Delta\delta^*$ ($''$)	$D^\dagger$ (kpc)	$W_0^\ddagger$ (Å)	$v^\S$ (km s^{-1})	S/N$^\|\|$
−22.0	−11.2	77.5	< 0.76	...	3.6
4.4	−10.4	40.3	< 0.76	...	3.6
3.6	−10.4	40.5	0.53 ± 0.14	55.7 ± 13.8	8.7
2.8	−10.4	41.3	< 0.59	...	4.6
−21.2	−10.4	76.5	< 0.60	...	4.5
−22.0	−10.4	78.0	< 0.79	...	3.4
4.4	−9.6	38.3	0.52 ± 0.18	15.7 ± 26.3	6.8
3.6	−9.6	38.5	< 0.26	...	10.6
2.8	−9.6	39.2	< 0.82	...	3.3
−21.2	−9.6	77.2	< 0.33	...	8.1
−22.0	−9.6	78.4	< 0.69	...	3.9
4.4	−8.8	36.2	< 0.49	...	5.5
3.6	−8.8	36.2	0.34 ± 0.12	50.3 ± 15.8	8.7
−21.2	−8.8	77.7	< 0.29	...	9.2
−22.0	−8.8	79.5	< 0.28	...	9.8
−22.8	−8.8	81.1	< 0.88	...	3.1
3.6	−8.0	33.8	0.73 ± 0.24	60.4 ± 25.0	5.1
2.8	−8.0	34.3	< 0.50	...	5.4
−21.2	−8.0	77.8	< 0.48	...	5.6
−22.0	−8.0	79.9	< 0.32	...	8.5
−22.8	−8.0	81.8	< 0.38	...	7.1
3.6	−7.2	31.2	< 0.79	...	3.5
2.8	−7.2	31.5	< 0.24	...	11.3
2.0	−7.2	32.6	< 0.40	...	6.8
−22.0	−7.2	80.1	< 0.65	...	4.2
−22.8	−7.2	82.4	< 0.45	...	6.1
2.8	−6.4	28.6	0.48 ± 0.17	34.2 ± 23.5	9.8
2.0	−6.4	29.5	< 0.23	...	12.1
1.2	−6.4	30.7	< 0.56	...	4.8
−22.8	−6.4	82.8	< 0.90	...	3.0
2.0	−5.6	26.3	0.47 ± 0.14	93.4 ± 25.9	7.4
1.2	−5.6	27.4	0.47 ± 0.10	51.7 ± 18.0	10.6
0.4	−5.6	28.9	< 0.56	...	4.9
1.2	−4.8	23.8	< 0.54	...	5.1
0.4	−4.8	25.2	0.53 ± 0.12	63.5 ± 12.3	7.2
−0.4	−4.8	27.1	1.24 ± 0.43	115.4 ± 29.0	3.2
0.4	−4.0	21.4	< 0.85	...	3.2
−0.4	−4.0	23.2	< 0.59	...	4.6
−1.2	−4.0	25.4	0.67 ± 0.26	83.7 ± 16.1	3.2
−1.2	−3.2	21.3	0.49 ± 0.11	75.3 ± 9.9	7.9
−2.0	−3.2	23.8	0.99 ± 0.15	90.7 ± 9.7	8.3
−1.2	−2.4	17.1	1.08 ± 0.19	87.5 ± 12.6	7.8
−2.0	−2.4	19.7	0.92 ± 0.11	51.6 ± 7.3	11.0
−2.8	−2.4	22.4	0.80 ± 0.24	31.4 ± 19.0	5.2
−3.6	−1.6	21.7	1.42 ± 0.20	36.4 ± 11.3	5.0
−4.4	−1.6	24.7	0.94 ± 0.14	56.1 ± 9.2	9.0
−5.2	−1.6	27.3	< 0.61	...	4.5
−9.2	−1.6	40.6	< 0.67	...	4.1
−5.2	−0.8	24.5	< 0.63	...	4.3
−6.0	−0.8	27.6	0.97 ± 0.26	67.2 ± 17.5	3.4

*Arc-position angular separation to G1 in the image plane.
†Projected physical separation to G1 in the absorber plane.
‡Mg II $\lambda = 2{,}796$ Å absorption strength (with 1σ error) or 2σ upper limit.
§Velocity relative to $z_{G1} = 0.98235$.
||Signal-to-noise ratio blueward of Mg II.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Upper limits on Mg II absorption near G2

$\Delta\alpha^*$ ($''$)	$\Delta\delta^*$ ($''$)	$D^\dagger$ (kpc)	$W_0^\ddagger$ (Å)	$v^\S$ (km s^{-1})	S/N$^\|\|$
−4.2	−1.6	22.0	< 0.81	...	3.4
−5.0	−1.6	25.2	< 0.70	...	3.9
−5.8	−1.6	28.9	< 0.32	...	8.6
−6.6	−1.6	32.5	< 0.41	...	6.7
−5.8	−0.8	27.1	< 0.79	...	3.4
−6.6	−0.8	30.9	< 0.60	...	4.5

*Arc-position angular separation to G2 in the image plane.
†Projected physical separation to G2 in the absorber plane.
‡Mg II $\lambda = 2{,}796$ Å absorption strength 2σ upper limit.
§Velocity relative to $z_{G1} = 0.98235$.
||Signal-to-noise ratio blueward of Mg II.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 3 | Absorption by Fe II and Mg I near G1

$\Delta\alpha$ ($''$)	$\Delta\delta$ ($''$)	D (kpc)	W_0^{2796} (Å)	W_0^{2382} (Å)	W_0^{2852} (Å)
−2.0	−3.2	23.8	0.99 ± 0.15	0.72 ± 0.22	< 0.35
−1.2	−2.4	17.1	1.08 ± 0.19	0.59 ± 0.21	0.44 ± 0.15
−4.4	−1.6	24.7	0.94 ± 0.14	0.62 ± 0.19	< 0.31

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 4 | Galaxy properties

ID	RA (deg)	DEC (deg)	z	v (km s^{-1})	$\Delta v_{\rm FWHM}$ (km s^{-1})	$m_{\rm F814W}$ (AB)	B-I
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G1−A	51.867229	−13.434300	0.98236	$+1.8 \pm 7.0$	138.0 ± 10.4	24.30 ± 0.03	0.68 ± 0.03
G1−B	51.867420	−13.434390	0.98267	$+48.8 \pm 12.0$	97.9 ± 18.4	24.64 ± 0.04	0.53 ± 0.04
G1−C	51.867229	−13.434060	0.98212	-35.2 ± 12.5	92.5 ± 17.9	24.85 ± 0.05	0.72 ± 0.05
G2	51.865139	−13.447130	0.98216	-29.2 ± 1.7	81.0 ± 2.8	24.27 ± 0.03	0.06 ± 0.03
G3	51.855511	−13.431970	0.98162	-109.8 ± 12.2	86.9 ± 17.4	23.99 ± 0.02	0.30 ± 0.02

ID	M_B	L/L^*	$\log M_*/{\rm M}_\odot$	$f_{\rm OII}$ (erg s^{-1} cm^{-2})	SFR (M$_\odot$ yr^{-1})	μ	$R_{\rm vir}$ (kpc)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G1−A	−18.37	0.05	9.5	$8.7 \cdot 10^{-18}$	0.23	2.6	92
G1−B	−17.92	0.03	9.1	$4.0 \cdot 10^{-18}$	0.11	2.7	79
G1−C	−17.81	0.03	9.1	$3.3 \cdot 10^{-18}$	0.09	2.5	79
G2	−18.86	0.08	9.0	$2.4 \cdot 10^{-17}$	0.85	2.0	73
G3	−19.08	0.10	9.4	$6.9 \cdot 10^{-18}$	0.28	1.7	85

Top row: columns (1)–(3) galaxy identification (ID) and coordinates (RA, right ascension; DEC, declination); (4) redshift (z); (5) velocity relative to $z_{\rm G1}=0.98235$ (v); (6) deconvolved [O II] line width ($\Delta v_{\rm FWHM}$); (7) apparent magnitude ($m_{\rm F814W}$); (8) rest-frame colour ($B-I$). Bottom row: column (2) absolute magnitude (M_B); (3) rest-frame B-band luminosity in terms of L^* at $z=0.98$ (L/L^*; ref. 48); (4) stellar mass ($\log(M_*/M_\odot)$; from SED fitting); (5) [O II] emission line flux ($f_{\rm OII}$); (6) star-formation rate (SFR; from emission line flux); (7) magnification (μ; subject to approximately 5% uncertainty); (8) virial radius ($R_{\rm vir}$). Magnification was used to correct quantities in columns (2)–(6) in the bottom row.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature25747

Multi-terminal memtransistors from polycrystalline monolayer molybdenum disulfide

Vinod K. Sangwan[1]*, Hong-Sub Lee[1]*, Hadallia Bergeron[1], Itamar Balla[1], Megan E. Beck[1], Kan-Sheng Chen[1] & Mark C. Hersam[1,2,3]

Memristors are two-terminal passive circuit elements that have been developed for use in non-volatile resistive random-access memory and may also be useful in neuromorphic computing[1–6]. Memristors have higher endurance and faster read/write times than flash memory[4,7,8] and can provide multi-bit data storage. However, although two-terminal memristors have demonstrated capacity for basic neural functions, synapses in the human brain outnumber neurons by more than a thousandfold, which implies that multi-terminal memristors are needed to perform complex functions such as heterosynaptic plasticity[3,9–13]. Previous attempts to move beyond two-terminal memristors, such as the three-terminal Widrow–Hoff memristor[14] and field-effect transistors with nanoionic gates[15] or floating gates[16], did not achieve memristive switching in the transistor[17]. Here we report the experimental realization of a multi-terminal hybrid memristor and transistor (that is, a memtransistor) using polycrystalline monolayer molybdenum disulfide (MoS_2) in a scalable fabrication process. The two-dimensional MoS_2 memtransistors show gate tunability in individual resistance states by four orders of magnitude, as well as large switching ratios, high cycling endurance and long-term retention of states. In addition to conventional neural learning behaviour of long-term potentiation/depression, six-terminal MoS_2 memtransistors have gate-tunable heterosynaptic functionality, which is not achievable using two-terminal memristors. For example, the conductance between a pair of floating electrodes (pre- and post-synaptic neurons) is varied by a factor of about ten by applying voltage pulses to modulatory terminals. *In situ* scanning probe microscopy, cryogenic charge transport measurements and device modelling reveal that the bias-induced motion of MoS_2 defects drives resistive switching by dynamically varying Schottky barrier heights. Overall, the seamless integration of a memristor and transistor into one multi-terminal device could enable complex neuromorphic learning and the study of the physics of defect kinetics in two-dimensional materials[18–22].

Uniform polycrystalline monolayer MoS_2 films with an average grain size of 3–5 μm were grown by chemical vapour deposition (CVD) on SiO_2/Si substrates (Methods) and characterized using X-ray photoelectron, photoluminescence and Raman spectroscopies (Fig. 1a–c and Extended Data Fig. 1). MoS_2 memtransistors were fabricated in a field-effect geometry with channel lengths (L) and widths (W) varying from 5 μm to 150 μm (Fig. 1d–f). Because clean interfaces between MoS_2 channels and metal electrodes were found to be critical, we developed an unconventional photolithography process based on a polymethylglutarimide (PMGI) and photoresist bilayer (Methods and Extended Data Fig. 2).

At large drain bias V_D in the sub-threshold regime (with gate bias, V_G, smaller than the threshold voltage, V_{th}), a typical memtransistor is in a high-resistance state (HRS) for the V_D sweep from 0 V to 80 V (sweep 1) and gradually changes to a low-resistance state (LRS) (Fig. 2a). The device maintains the LRS during the sweep from 80 V to 0 V (sweep 2), is reset to the HRS during the sweep from 0 V to −80 V (sweep 3), and maintains the HRS from −80 V to 0 V (sweep 4). Thus, these devices act as LRS–HRS memtransistors. Figure 2b shows that when V_G is varied from 50 V to −50 V, both the LRS and HRS resistances change by a factor of about 10^4 and the switching ratio (I_{LRS}/I_{HRS} at $V_D = 0.5$ V) is reduced from 300 to 8 (Fig. 2e inset). Owing to the n-type MoS_2 channel, the forward-biased ($V_D > 0$ V) device is completely off at $V_G = -50$ V (Fig. 2c), and the reverse-biased ($V_D < 0$ V) device is insulating for a range of V_D values, depending on the applied V_G (Fig. 2b and Extended Data Fig. 3c). The gate leakage current (I_G) remains below 200 pA during high-V_D and -V_G sweeps (Extended Data Fig. 3a).

Unlike filament-based resistive switching, MoS_2 memtransistors do not require an electroforming process to train the device, although the switching ratio increases with increasing range of the V_D sweep (Extended Data Fig. 3b). The largest switching ratio, higher than 100, was obtained from devices with $W = 100$–150 μm and $L = 5$–15 μm. These devices showed bipolar resistive switching, where reversing the bias polarity is essential to restoring the initial resistance states (Extended Data Fig. 3d, e). The hysteresis in the I_D–V_D curves of these MoS_2 memtransistors is fundamentally different from the commonly reported hysteresis in the transfer characteristics (I_D–V_G curves) of field-effect transistors (FETs), which are typically due to oxide-related traps[23,24]. Instead, the I_D–V_G curves of MoS_2 memtransistors in the LRS and the HRS show large shifts (about 10 V) in threshold voltage, and intersect at $V_G = V_{cross}$ (Fig. 2c). Compared to the LRS, the HRS shows up to 100 times higher resistance for $V_G < V_{cross}$ and 2 times higher field-effect mobility (μ) at $V_G > V_{cross}$. Therefore, the forward-bias switching loop changes from anticlockwise (LRS to HRS) for $V_G < V_{cross}$ to clockwise (HRS to LRS) for $V_G > V_{cross}$ (Fig. 2b and Extended Data Fig. 3f).

Figure 2d shows the endurance characteristics of a MoS_2 memtransistor that was switched 475 times between the LRS and the HRS using full-sweep cycles. Within a subset of these cycles, I_D saturates at an upper value via stretched exponentials (Extended Data Figs 3g–k, 4a). Between neighbouring subsets, I_D jumps randomly to a value about 10 times smaller, followed by the same inverse exponential growth, which suggests an oxide-related trap–release process activated by large fields near the source electrode in the forward bias[23]. This behaviour is reduced under reverse bias (Extended Data Fig. 4b), possibly owing to smaller band-bending near the drain electrode. Because the HRS and the LRS show similar transitory decays, the switching ratio (about 100) remains relatively constant (Fig. 2d). Individual resistance states measured within periods of up to 24 h show a projected retention of distinct states for timescales of the order of years (Fig. 2e and Extended Data Fig. 4e). A statistical study of 62 devices fabricated with identical geometry on a single chip showed a logarithmic normal distribution

[1]Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, USA. [2]Department of Chemistry, Northwestern University, Evanston, Illinois 60208, USA. [3]Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois 60208, USA.
*These authors contributed equally to this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

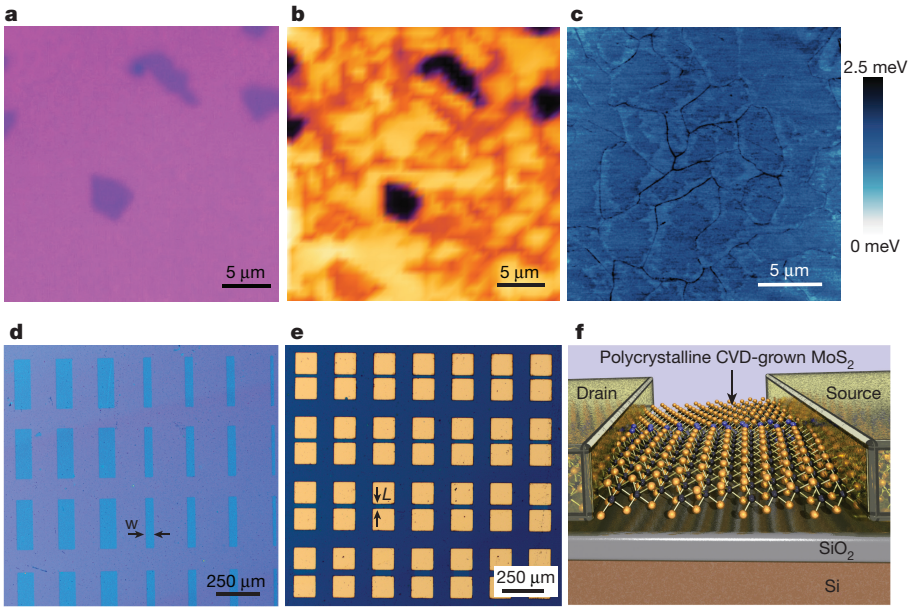


Figure 1 | Architecture of the MoS_2 memtransistor. a, Optical micrograph of CVD-grown polycrystalline monolayer MoS_2. The darker regions indicate sparse growth of bilayer MoS_2. **b**, Spatial mapping of photoluminescence intensity (wavelength, 670 nm) for the area shown in **a**. The darkest regions correspond to bilayer MoS_2, while the red features correspond to grain boundaries within monolayer MoS_2. **c**, Lateral force microscopy retrace image of monolayer MoS_2, showing grain boundaries (see Extended Data Fig. 1e for topography image). The colour bar shows the photodiode readout corresponding to the cantilever tilt. **d**, Optical micrograph of an array of MoS_2 monolayer strips with varying width, W, etched by reactive ion etching before metallization. **e**, Optical micrograph of fabricated MoS_2 memtransistors with varying channel length, L. **f**, Schematic of a MoS_2 memtransistor device built on 300-nm-thick thermal SiO_2 on doped Si (gate).

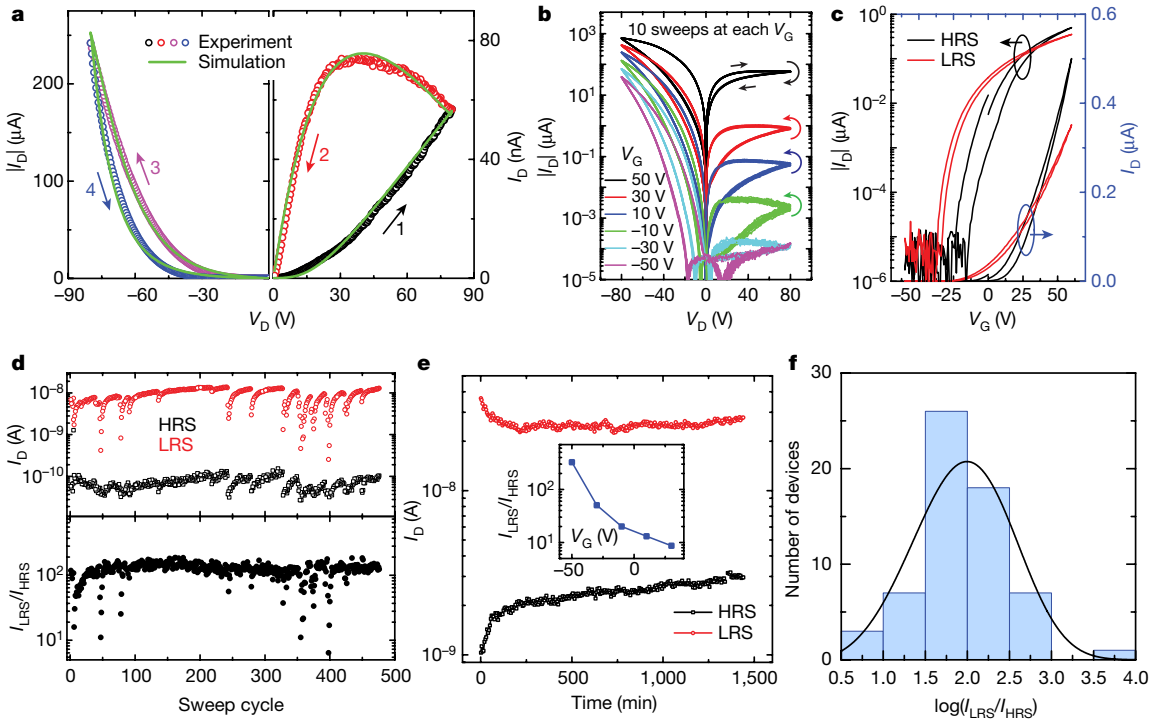


Figure 2 | Electrical characteristics of MoS_2 memtransistors. a, I_D–V_D curve (open circles) of a MoS_2 memtransistor ($L = 5\,\mu m$, $W = 100\,\mu m$) at gate bias $V_G = 10$ V. The direction of the drain bias (V_D) sweep $1 \rightarrow 4$ is indicated by coloured arrows. The sweep rate is 10 V s^{-1}throughout. The solid green line represents simulated data from a memtransistor model (see Methods and Extended Data Fig. 7). **b**, I_D–V_D curves for ten consecutive sweeps at each gate bias V_G for the same device. The switching directions are shown by the curved arrows. V_G was decreased for each consecutive sweep cycle. **c**, Transfer characteristics of a memtransistor at $V_D = 0.1$ V, showing a shift in threshold voltage and field-effect mobility from $V_{th} = 20$ V and $\mu \approx 0.6\,cm^2\,V^{-1}\,s^{-1}$ in the HRS to $V_{th} = 10$ V and $\mu \approx 0.3\,cm^2\,V^{-1}\,s^{-1}$ in LRS. The curves intersect at $V_G = V_{cross} \approx 30$ V. **d**, Endurance of the current (top) and I_{LRS}/I_{HRS} (bottom) at $V_D = 0.5$ V in the HRS (sweep 1) and the LRS (sweep 2) for 475 consecutive sweep cycles of a memristor at $V_G = 40$ V (see Extended Data Fig. 3d–k for all I_D–V_D curves). **e**, Retention of the HRS and LRS currents at $V_D = 100$ mV and $V_G = 0$ V in a 24-h period. The inset shows the gate tunability of I_{LRS}/I_{HRS}. **f**, Histogram showing the largest I_{LRS}/I_{HRS} ratios of 62 distinct memtransistors ($L = 5\,\mu m$, $W = 100\,\mu m$). The distribution is fitted with a logarithmic normal curve with mean and variance (in $\log(I_{LRS}/I_{HRS})$) of 1.96 and 0.54, respectively.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

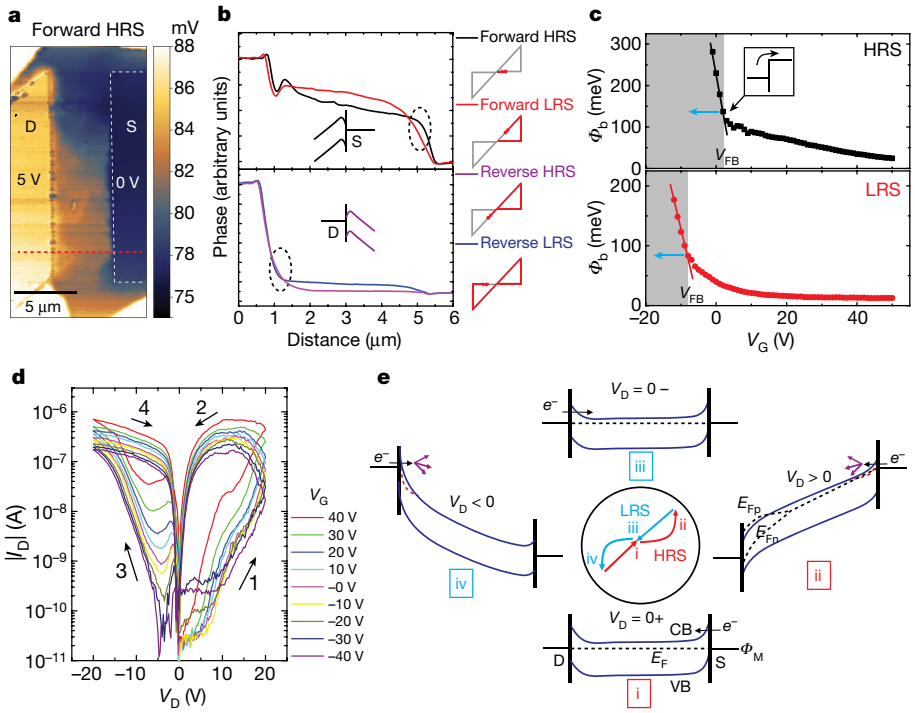


Figure 3 | *In situ* measurements and switching mechanism. a, EFM phase image of a MoS_2 memtransistor at $V_D = 5$ V, $V_S = 0$ V and $V_{tip} = 0$ V (V_{tip}, voltage of the EFM tip) in the forward-biased HRS. D, drain; S, source. **b**, Line profiles of EFM phase along the red dashed line in **a** for EFM images taken in the forward-biased HRS, forward-biased LRS, reverse-biased HRS and reverse-biased LRS, respectively, as shown schematically on the right (see Extended Data Fig. 5c–f for EFM images). The dashed-line ovals show band bending near the source electrode in forward-biased HRS and LRS, and near the drain in the reverse-biased LRS and HRS. The dip that appears near the drain in the top panel is a topography artefact. **c**, Dependence of the effective Schottky barrier height (Φ_b) on V_G, extracted from variable-temperature conductivity measurements in the HRS (top) and the LRS (bottom). For $V_G = V_{FB}$, Φ_b deviates from the linear dependence on V_G of the thermionic emission model. The inset shows the flat-band condition at $V_G = V_{FB} = 2$ V, $\Phi_b = 125$ meV in the HRS and $V_{FB} = -8$ V, $\Phi_b = 80$ meV in the LRS (see Methods and Extended Data Fig. 6). **d**, I_D–V_D curve of an LRS–LRS MoS_2 memtransistor with an approximately 1.5-nm-thick photoresist layer under the metal contact at different V_G values. The direction of the V_D sweep $1 \rightarrow 4$ is indicated by arrows. V_G was decreased between V_D sweep cycles. **e**, Schematic showing the energy band diagram of an LRS–LRS memtransistor at the four switching stages shown in the centre. E_{Fp} and E_{Fn} are non-equilibrium quasi-Fermi levels for holes and electrons, respectively (see Methods and Extended Data Fig. 8 for details). CB and VB stand for conduction band and valence band, respectively. Purple arrows show defect migration. See Methods section 'The switching mechanism' for detailed description of stages i-iv.

of switching ratios (Fig. 2f). Device-to-device variability is attributed to spatial inhomogeneity in the CVD-grown MoS_2 film.

Owing to its two-dimensional (2D) nature, the MoS_2 channel allows the switching mechanism to be probed via *in situ* electrostatic force microscopy (EFM), as shown in Fig. 3a and b. The drain contact and conductive cantilever tip were biased independently (Extended Data Fig. 5), while the phase shift of the cantilever was recorded as the tip traced the topography profile of the MoS_2 channel and electrodes 50 nm above the surface. Because the phase shift is proportional to the square of the potential difference between the tip and the surface, the phase image provides a map of the local potential[20]. Line profiles reveal a sharper potential drop (larger field) at the source in the forward-biased HRS compared to the forward-biased LRS, suggesting a larger contact resistance in the HRS than in the LRS (Fig. 3b). The reverse-biased HRS also shows a larger field at the drain compared to the reverse-biased LRS, although this difference is smaller, which is consistent with the smaller switching ratio at reverse bias. Multiple line profiles confirm that the differences between the HRS and LRS electric fields are consistent along the entire channel width, despite grain-boundary-induced EFM phase variations far from the contacts (Extended Data Fig. 5g–o). Because a reverse-biased Schottky diode at the source (drain) dominates I_D at $V_D > 0$ V ($V_D < 0$ V), EFM measurements provide direct evidence that the switching from the HRS to the LRS (from the LRS to the HRS) in memtransistors is caused by the dynamic tuning of Schottky barriers. Using charge-transport measurements with variable temperature (T) and assuming a 2D thermionic emission model, the effective barrier heights (Φ_b) in different states are extracted from the slope of $\ln(I_D/T^{3/2})$ versus $1/T$ at different V_G values (Extended Data Fig. 6). Thermionic emission dominates at $V_G < V_{FB}$ (flat-band voltage), producing Φ_b values that vary linearly with V_G. Thermally assisted tunnelling through a deformed Schottky barrier begins to contribute to I_D at $V_G > V_{FB}$, resulting in deviation from the linear trend (Fig. 3c). Thus, the Schottky barrier height ($\Phi_{sb} = \Phi_b$) is extracted from the relationship $V_G = V_{FB}$. V_{FB} decreases from 2 V (HRS) to −8 V (LRS), which is consistent with a V_{th} shift of 15 V for the same device (Extended Data Fig. 6). In addition, Φ_{sb} decreases from 125 meV (HRS) to 80 meV (LRS), confirming the EFM observations.

Charge transport in MoS_2 LRS–HRS memtransistors can be described by Schottky barrier transistors (equations (1) and (2), where Φ_b is a function of an internal state variable (w) defined as the width of the region with excess dopants (Δn) (Extended Data Fig. 7a)[1,17,25]. Tuning the Schottky barrier through increased doping near contacts ($\Phi_b \approx \sqrt{w\Delta n}$) is a standard practice in conventional FETs[26,27]. Similarly, we propose that defects in MoS_2 act as dopants and their local migration under an applied bias is facilitated by grain boundaries, as previously observed using transmission electron microscopy and explained by *ab initio* calculations[18,19,21]. Thus, we develop a memtransistor model in which Φ_b changes by image charge lowering and tunnelling at high biases (see Methods), resulting in the following coupled equations:

$$I_D = D\exp\left[\frac{e(V_G - V_{th})}{c_r k_B T}\right]\left[1 - \exp\left(-\frac{eV_D}{k_B T}\right)\right]\exp\left[\frac{\Phi_b(w)}{k_B T}\right] \quad (1)$$

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

$$\Phi_b(w) = \phi_{b0} - \frac{e}{\varepsilon_s}\sqrt{\frac{w\Delta n}{4\pi}} + \sqrt{\frac{e}{4\pi\varepsilon_s}}\left[\frac{2en(\phi_{b0} + A|V_D|)}{\varepsilon_s}\right]^{1/4} \quad (2)$$

$$\frac{\partial w}{\partial t} = EI_D\{1 - [(w - 0.5)^2 + 0.75]^p\} \quad (3)$$

where A, D, E and ϕ_{b0} are fitting parameters, and e, n, ε_s and k_B are the charge of the electron, the doping level, the dielectric constant of monolayer MoS_2 and the Boltzmann constant, respectively. Equation (3) describes the nonlinear kinetics of the dopants near the contacts by using a window function, where the degree of nonlinearity is defined by a positive integer p ($p = 4$ in Fig. 2a; see Methods)[28]. This model agrees well with the experimental data (Fig. 2a and Extended Data Fig. 7) and helps establish that the dependence of w on V_G (through I_D) is the most important feature of the MoS_2 memtransistor, distinguishing it from a two-terminal memristor[17]. To further illustrate this non-trivial switching mode, we fabricated MoS_2 devices without the PMGI-based process and used a residual photoresist layer about 1.5 nm thick as a tunnelling barrier between MoS_2 and the metal contacts (Fig. 3d, e and Extended Data Fig. 8). The tunnelling barrier minimizes pinning of the Fermi level and causes the resistance states to change abruptly upon crossing the 0 V level, resulting in LRS–LRS memtransistors. Overall, the switching mechanism of both LRS–HRS and LRS–LRS memtransistors can be described by two memristors at the contacts connected by a FET (see Methods section 'The switching mechanism' and Extended Data Fig. 8e).

While grain boundaries in polycrystalline MoS_2 memtransistors enable large switching ratios and prevent electrical breakdown by lowering Schottky barriers through dynamic defect migration, control devices without grain boundaries on single grains of CVD-grown MoS_2 show a qualitatively different, reversible breakdown phenomenon (Fig. 4a). This breakdown is marked by a sharp drop in the conductance at a voltage V_{br} during sweep 1, with a subsequent increase to the original conductance value during sweep 2. This abrupt change in charge transport is accompanied by the emergence of dendritic features (about 300–500 nm in length) close to the source electrode (upper inset in Fig. 4a and Extended Data Fig. 9a, b). Control devices reveal a linear correlation of the breakdown current (I_{br}) with the width of the source electrode (W_s) and no correlation with L (lower inset in Fig. 4a and Extended Data Fig. 9f–j). Because I_D is normally proportional to W/L, the observed deviation from this relationship at breakdown suggests that the reverse-biased Schottky diode at the source electrode creates a bottleneck for electron injection at the channel[27], ultimately causing electromigration at the source contact[7]. This reversible breakdown in CVD-grown MoS_2 differs from the irreversible breakdown induced in exfoliated MoS_2 by Joule heating[29]. In the high-bias limit (±120 V), polycrystalline MoS_2 memtransistors also degrade irreversibly in a manner that shows light emission in each subsequent sweep (Extended Data Fig. 9c–e).

The 2D planar geometry of the MoS_2 memtransistor allows the realization of multi-terminal neural circuits that mimic multiple synaptic connections in neurons. For example, in a six-terminal memtransistor, the conductance between any two of the four inner electrodes can be modulated by high-bias pulses applied to the two outer electrodes while the inner electrodes are disconnected (Fig. 4b and Extended Data Fig. 10). To achieve heterosynaptic plasticity, the conductance between pre-synaptic and post-synaptic neurons should be controlled by additional modulatory terminals[30]. While this type of modulation has been demonstrated previously in Ag-based cationic memristors, this design is limited to only three terminals, owing to the requirement of filament formation across the channel[6,30]. In contrast, the MoS_2 memtransistor can have a larger number of terminals and allows facile tuning through the modulation of the local Schottky barrier at each terminal (see Methods section 'The switching mechanism'). Multi-terminal memtransistors also allow further tuning of heterosynaptic plasticity through a gate electrode, in which the switching ratio for any pair of side electrodes can be increased by about two to ten times by varying V_G from 50 V to 20 V (Extended Data Fig. 10f).

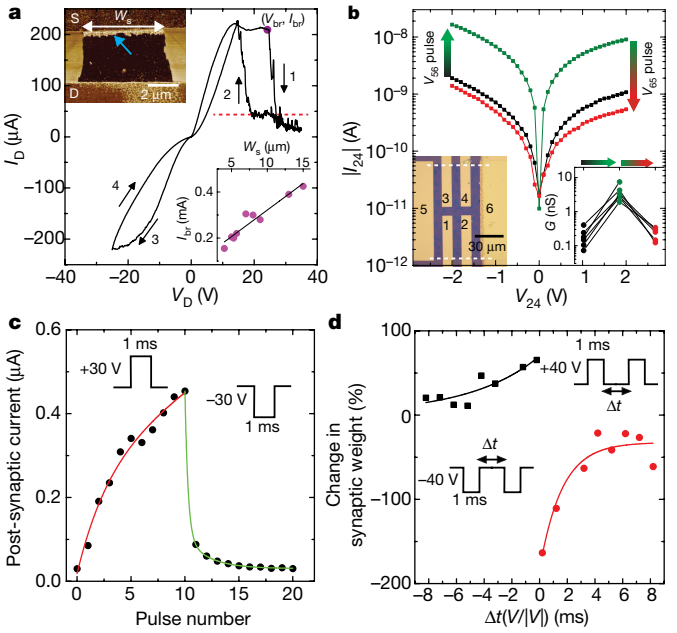


Figure 4 | Control devices and neural functions of memtransistors. **a**, I_D–V_D curve (at $V_G = 60$ V) of a control device on an individual monolayer MoS_2 domain without grain boundaries, showing breakdown. The breakdown voltage V_{br} and breakdown current I_{br} are defined by the point preceding the sharp downward transition in sweep 1. The upper inset shows an atomic force microscopy phase image of a device with degraded material (arrow) near the source electrode. The lower inset shows the linear relation between I_{br} and the width of the source contact, W_s, for nine devices with different flake shapes (see Extended Data Fig. 9). **b**, I_{24}–V_{24} curve between terminals 2 and 4 of a six-terminal MoS_2 memtransistor (left inset) at constant $V_G = 20$ V, at the following three stages: (1) before any pulse (black); (2) after applying four pulses of −80 V between terminals 5 and 6 while terminals 1–4 were disconnected (green); and (3) after applying three pulses of −80 V between terminals 6 (drain) and 5 (ground) while terminals 1–4 were disconnected (red). The white dashed lines in the left inset show the edges of the patterned MoS_2. The right inset shows heterosynaptic plasticity, observed by reversibly changing the conductance ($G = I_{ij}/V_{ij}$) between the six combinations of four terminals (ij = 12, 13, 14, 23, 24 and 34) by a factor of about 10 (see Extended Data Fig. 10). **c**, Post-synaptic current versus pulse number, showing long-term potentiation and depression with 30-V and −30-V pulses. The solid lines are biexponential fits. **d**, Measured change in synaptic weight (normalized to maximum weight) as a function of the time interval (Δt) between paired pulses of 40 V and −40 V. The solid lines are exponential fits with time constants of 1.6 ms (red) and 5.5 ms (black) for positive and negative pulses, respectively. The inset shows the timing scheme for indirect spike-timing-dependent plasticity. $V_G = 0$ V for all measurements in **c** and **d**.

MoS_2 memtransistors also demonstrate long-term potentiation and depression, which mimic excitatory and inhibitory synapses in organisms (Fig. 4c). The post-synaptic current is shown to increase and decrease exponentially with the repetition of positive- and negative-bias pulses of 1 ms (Fig. 4c). The linearity of the pulse train is comparable to that of metal-oxide memristors and can be further improved by employing bipolar pulsing schemes[11,31]. By mimicking indirect spike-timing-dependent plasticity, paired pulses separated by a time interval induce positive and negative changes in the synaptic weight using positive and negative pulses (Fig. 4d), resulting in time constants of about 2 ms and 6 ms, respectively, which are comparable to the response times of biological synapses[11,32].

In conclusion, MoS_2 memtransistors combine resistive switching with transistor gating to realize nonlinear charge transport with wide

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

tunability of individual states and switching ratios. In contrast to conventional devices that require single-crystal MoS_2 flakes, the utilization of polycrystalline and all-surface MoS_2 films allows the straightforward scaling of this technology to large-area integrated circuits and post-growth defect engineering. The 2D planar geometry of monolayer MoS_2 further enables the realization of multi-terminal memtransistors with unprecedented heterosynaptic plasticity. This technology may enable complex learning from multiple inputs in neuromorphic computing by mimicking biological neurons with multiple synapses.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 June; accepted 19 December 2017.

1. Strukov, D. B., Snider, G. S., Stewart, D. R. & Williams, R. S. The missing memristor found. *Nature* **453,** 80–83 (2008); corrigendum **459,** 1154 (2009).
2. Yang, J. J. *et al.* Memristive switching mechanism for metal/oxide/metal nanodevices. *Nat. Nanotechnol.* **3,** 429–433 (2008).
3. Kuzum, D., Yu, S. & Wong, H. S. P. Synaptic electronics: materials, devices and applications. *Nanotechnology* **24,** 382001 (2013).
4. Yu, S. & Chen, P. Y. Emerging memory technologies: recent trends and prospects. *IEEE Sol. Stat. Circuit Mag.* **8,** 43–56 (2016).
5. Yang, J. J., Strukov, D. B. & Stewart, D. R. Memristive devices for computing. *Nat. Nanotechnol.* **8,** 13–24 (2013).
6. Waser, R., Dittmann, R., Staikov, G. & Szot, K. Redox-based resistive switching memories – nanoionic mechanisms, prospects, and challenges. *Adv. Mater.* **21,** 2632–2663 (2009).
7. Pershin, Y. V. & Di Ventra, M. Memory effects in complex materials and nanoscale systems. *Adv. Phys.* **60,** 145–227 (2011).
8. Wong, H. S. P. *et al.* Metal-oxide RRAM. *Proc. IEEE* **100,** 1951–1970 (2012).
9. Thomas, A. Memristor-based neural networks. *J. Phys. D* **46,** 093001 (2013).
10. Indiveri, G. *et al.* Integration of nanoscale memristor synapses in neuromorphic computing architectures. *Nanotechnology* **24,** 384010 (2013).
11. Jo, S. H. *et al.* Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett.* **10,** 1297–1301 (2010).
12. Kim, S. *et al.* Pattern recognition using carbon nanotube synaptic transistors with an adjustable weight update protocol. *ACS Nano* **11,** 2814–2822 (2017).
13. Kim, S. *et al.* Experimental demonstration of a second-order memristor and its ability to biorealistically implement synaptic plasticity. *Nano Lett.* **15,** 2203–2211 (2015).
14. Widrow, B. *An Adaptive Adaline Neuron using Chemical Memristors.* Technical Report 1553–2 www-isl.stanford.edu/~widrow/papers/t1960anadaptive.pdf (Stanford Electronics Laboratories, 1960).
15. Lai, Q. *et al.* Ionic/electronic hybrid materials integrated in a synaptic transistor with signal processing and learning functions. *Adv. Mater.* **22,** 2448–2453 (2010).
16. Diorio, C., Hasler, P., Minch, A. & Mead, C. A. A single-transistor silicon synapse. *IEEE Trans. Electron Dev.* **43,** 1972–1980 (1996).
17. Mouttet, B. Memristive systems analysis of 3-terminal devices. In *IEEE Int. Conf. on Electronics, Circuits and Systems* 930–933 (IEEE, 2010).
18. Azizi, A. *et al.* Dislocation motion and grain boundary migration in two-dimensional tungsten disulphide. *Nat. Commun.* **5,** 4867 (2014).
19. Komsa, H.-P. *et al.* From point to extended defects in two-dimensional MoS_2: Evolution of atomic structure under electron irradiation. *Phys. Rev. B* **88,** 035301 (2013).
20. Sangwan, V. K. *et al.* Gate-tunable memristive phenomena mediated by grain boundaries in single-layer MoS_2. *Nat. Nanotechnol.* **10,** 403–406 (2015).
21. Yu, Z. G., Zhang, Y.-W. & Yakobson, B. I. An anomalous formation pathway for dislocation-sulfur vacancy complexes in polycrystalline monolayer MoS_2. *Nano Lett.* **15,** 6855–6861 (2015).
22. Bessonov, A. A. *et al.* Layered memristive and memcapacitive switches for printable electronics. *Nat. Mater.* **14,** 199–204 (2015).
23. He, G. *et al.* Thermally assisted nonvolatile memory in monolayer MoS_2 transistors. *Nano Lett.* **16,** 6445–6451 (2016).
24. Arnold, A. J. *et al.* Mimicking neurotransmitter release in chemical synapses via hysteresis engineering in MoS_2 transistors. *ACS Nano* **11,** 3110–3118 (2017).
25. Chua, L. Resistance switching memories are memristors. *Appl. Phys., A Mater. Sci. Process.* **102,** 765–783 (2011).
26. Shannon, J. M. Control of Schottky barrier height using highly doped surface layers. *Solid-State Electron.* **19,** 537–543 (1976).
27. Sze, S. M. & Ng, K. K. *Physics of Semiconductor Devices* (Wiley-Interscience, 2006).
28. Prodromakis, T., Peh, B. P., Papavassiliou, C. & Toumazou, C. A versatile memristor model with nonlinear dopant kinetics. *IEEE Trans. Electron Dev.* **58,** 3099–3105 (2011).
29. Lembke, D. & Kis, A. Breakdown of high-performance monolayer MoS_2 transistors. *ACS Nano* **6,** 10070–10075 (2012).
30. Yang, Y., Chen, B. & Lu, W. D. Memristive physically evolving networks enabling the emulation of heterosynaptic plasticity. *Adv. Mater.* **27,** 7720–7727 (2015).
31. Wang, I. T. *et al.* 3D Ta/TaO_x/TiO_2/Ti synaptic array and linearity tuning of weight update for hardware neural network applications. *Nanotechnology* **27,** 365204 (2016).
32. Bi, G.-Q. & Poo, M.-M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18,** 10464–10472 (1998).

Acknowledgements This research was supported by the Materials Research Science and Engineering Center (MRSEC) of Northwestern University (NSF DMR-1720139) and the 2-DARE programme (NSF EFRI-1433510). The CVD growth of MoS_2 was supported by the National Institute of Standards and Technology (NIST CHiMaD 70NANB14H012). Charge transport instrumentation was funded by an ONR DURIP grant (ONR N00014-16-1-3179). H.-S.L. acknowledges the Basic Science Research Program of the National Research Foundation of Korea (NRF), which is funded by the Ministry of Education (2017R1A6A3A03008332). H.B. acknowledges support from the NSERC Postgraduate Scholarship–Doctoral Program. H.B. and M.E.B. acknowledge support from the National Science Foundation through a Graduate Research Fellowship. For this work, we used the Northwestern University NUANCE Center and the Northwestern University Micro/Nano Fabrication Facility (NUFAB), which are partially supported by the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-1542205), the Materials Research Science and Engineering Center (NSF DMR-1720139), the State of Illinois and Northwestern University. We thank J. J. McMorrow for assistance with photolithography, X. Liu for assistance with lateral force microscopy and S. Mohseni for assistance with atomic force microscopy.

Author Contributions V.K.S., H.-S.L. and M.C.H. conceived the idea and designed the experiments. V.K.S. and H.-S.L. fabricated all the devices and performed measurements and analysis. H.B. and I.B. handled the growth of MoS_2 and conducted materials characterization. V.K.S. developed the memtransistor model. M.E.B. assisted in model fitting and device fabrication. K.-S.C. assisted in electrostatic force microscopy. All authors wrote the manuscript and discussed the results at all stages.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.C.H. (m-hersam@northwestern.edu).

Reviewer Information *Nature* thanks X. Liang and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

CVD of MoS_2 film. Polycrystalline monolayer MoS_2 with continuous film coverage was directly grown on oxidized Si substrates by CVD using sulfur powder (Sigma-Aldrich) and molybdenum trioxide powder (MoO_3, 99.98% trace metal, Sigma-Aldrich) following the procedure reported earlier[33]. Prior to growth, Si substrates were bath-sonicated for 10 min in acetone and isopropyl alcohol and subsequently cleaned under O_2 plasma (Harrick Plasma) at about 200 mTorr for 2 min with 10.2 W power applied to the radiofrequency coil. The substrates were placed downstream of 12 mg of MoO_3 powder in an alumina boat positioned in a 1-inch-diameter quartz tube furnace (Lindberg/Blue). 150 mg of sulfur powder in an alumina boat was placed about 30 cm upstream of the MoO_3 boat (outside the furnace) and was heated independently using a temperature-controlled heating belt. The tube furnace was purged using ultrahigh-purity Ar gas at 200 standard cubic centimetres per minute (sccm) for 10 min. Then, the pressure was increased to 400 Torr, and the tube was evacuated to its base pressure of about 60 mTorr. The purging process was repeated twice to achieve an inert growth environment. The pressure was kept at 150 Torr with a 25-sccm flow of Ar gas during growth and cooling. After the purge process, the furnace was heated to 150 °C for 5 min and held at that temperature for 20 min to remove physisorbed contaminants from the growth environment. The furnace was then heated to 800 °C at a rate of 12 °C min^{-1} and held at that temperature for 20 min, followed by natural cooling to room temperature. Concurrently, the heating belt around the sulfur boat was first ramped to 50 °C for 5 min and was held at that temperature for 49 min. The heating belt was then brought to 150 °C at a rate of 4.5 °C min^{-1} and held at that temperature for an additional 23 min, after which it was allowed to cool to room temperature naturally.

Chemical characterization of monolayer MoS_2 film. The coverage and growth quality of continuous polycrystalline monolayer CVD-grown MoS_2 on 5 mm × 5 mm substrates were characterized by optical microscopy, Raman microscopy and photoluminescence spectroscopy. Raman spectra collected with a Horiba Scientific XploRA PLUS Raman microscope (Extended Data Fig. 1a) show a peak spacing of about 19 cm^{-1} between the in-plane A_{1g} and the out-of-plane E^1_{2g} breathing modes, thus confirming[34] the growth of monolayer MoS_2. Photoluminescence spectroscopy (Extended Data Fig. 1b) shows a dominant exciton A at 673 nm, typical of CVD-grown monolayer MoS_2 on SiO_2.

The chemical composition of CVD-grown polycrystalline monolayer MoS_2 films was confirmed by X-ray photoelectron spectroscopy (XPS) measurements using a Thermo Scientific ESCA Laboratory 250Xi scanning XPS instrument connected with a monochromatic Kα Al X-ray line. The X-ray beam size was approximately 900 μm in diameter with an elliptical cross-section. A charge-neutralization Ar^+ ion flood gun was also used to compensate for local electrostatic fields arising from charge buildup in the MoS_2/SiO_2 samples. All XPS spectra were analysed using Avantage software (Thermo Scientific). All Mo 3*d* subpeaks were fitted using floating Gaussian–Lorentzian mixing and modified Shirley background subtraction. Fitting of Mo 3*d* subpeaks ensured that their full-width at half-maximum values were lower than 3 eV, and doublets were also constrained to have the same full-width at half-maximum. In the XPS spectra analysis, charging effects were compensated by correcting all spectra against the C 1*s* adventitious carbon peak at 284.8 eV. As shown in Extended Data Fig. 1c and d, the XPS spectra of the Mo 3*d* and S 2*p* orbitals are consistent with previous reports[35] of stoichiometric CVD-grown MoS_2 with a small amount of MoO_x in the MoS_2 film[33].

Scanning probe microscopy. The CVD-grown MoS_2 films were characterized by atomic force microscopy (AFM) in ambient conditions using an Asylum Cypher AFM system in lateral force microscopy (LFM) mode and tapping mode. For LFM, NanoWorld FMR cantilevers with a resonance frequency of about 75 kHz and contact force of approximately 5 nN were used (Fig. 1c). The LFM retrace image was acquired by scanning the tip from right to left with a contact force of about 5 nN and scan rate of 1 Hz. For the tapping mode, NanoWorld NCHR-W Si cantilevers (resonance frequency of about 320 kHz) were used (Extended Data Fig. 1e, f). A step height of 0.73 nm at the edge of a MoS_2 flake confirmed monolayer thickness. AFM measurements of individual flake devices after electrical breakdown (upper inset in Fig. 4a and Extended Data Fig. 9) were conducted in the tapping mode using a Bruker Dimension FastScan AFM system. EFM of MoS_2 memtransistors was conducted in ambient conditions using the environment cell of the Asylum Cypher AFM system and NanoWorld PointProbe EFM tips coated with PtIr (resonance frequency of about 75 kHz).

Fabrication of MoS_2 memtransistors. MoS_2 LRS–HRS memtransistors were fabricated following a photolithography process developed to minimize processing residues. First, a PMGI film was spin-coated on SiO_2/Si substrates covered by polycrystalline monolayer MoS_2 at 4500 r.p.m. for 45 s, followed by baking at 170 °C for 10 min using a hot plate. Note that the baking conditions for the PMGI layer were found to be critical. Lower temperature (160 °C) resulted in delamination during development, while higher temperature (180 °C) made it difficult to remove PMGI during the lift-off step. Afterwards, a thin film of photoresist (S1813, Shipley Company) was spin-coated at 4000 r.p.m. for 1 min followed by another pre-baking step at 115 °C for 1 min. The bilayer resist was exposed to ultraviolet light (365 nm) for 15 s and then developed in MF319 developer (Shipley Company) for 20 s without post-baking. The samples were rinsed with deionized water and thoroughly dried before subjecting them to reactive ion etching to pattern the exposed MoS_2 film. The MoS_2 film was etched using Ar at a power of 50 W, pressure of 100 mTorr and flow rate of 50 sccm for 20 s. Subsequently, the bilayer resist mask was removed by submerging the substrate overnight in an N-methyl-2-pyrrolidone (NMP) bath heated at 80 °C. At this stage, the patterned MoS_2 strips resembled the optical image shown in Fig. 1e. Finally, source and drain electrodes were fabricated by patterning a negative photoresist NR9-1000PY (Futurrex, Inc.), followed by thermal evaporation of metals (3 nm Ti and 50 nm Au) and lift-off in NMP. The MoS_2 LRS–LRS memtransistors were fabricated using the same process, but without a PMGI layer.

Electrical measurements. All room-temperature electrical measurements were carried out in a vacuum probe station at a pressure of about 5×10^{-5} Torr in the dark using a LakeShore CRX 4K probe station. Endurance, retention and neural learning tests were conducted in vacuum using pulse measure units in a Keithley 4200A-SCS Parameter Analyser and home-built LabVIEW programs. For the variable-temperature transport measurements, the devices were first switched to the HRS, and the I_D–V_G data were collected at $V_D = 0.1$ V from 300 K to 75 K at a step of 25 K. A pressure of 8×10^{-7} Torr was achieved during cryogenic measurements using a built-in cryopump. The devices were allowed to heat up to room temperature overnight and were then switched to the LRS at room temperature, followed by similar variable-temperature transport measurements in the LRS. In the sub-threshold regime, the charge transport in Schottky-barrier FETs (SB-FETs) is dominated by thermionic emission:[27,36]

$$I_D = A^* T^{3/2} \exp\left(\frac{\Phi_b}{k_B T}\right)\left[\exp\left(\frac{eV_D}{k_B T}\right) - 1\right] \quad (4)$$

where A^* is the 2D equivalent Richardson constant. The term $T^{3/2}$ comes from the 2D model (as opposed to T^2 in three dimensions). At low bias, we can extract Φ_b from the relationship of $\ln(I/T^{3/2})$ with $1/T$ at different V_G values (Extended Data Fig. 6).

Modelling the MoS_2 memtransistor. We model the memtransistor behaviour by integrating a mathematical formalism of memristive systems[1,2,25,36,37] and memory transistors (memtransistors)[17] with the charge transport model of an SB-FET[26,27,38–41]. Memristive systems are defined as:

$$\frac{dw}{dt} = f(w, V, t) \text{ and } I = g(w, V, t)V \quad (5)$$

where t is the time, and V and I are the input (voltage) and output (current) of the system, respectively. The internal state variable w is represented by an n-dimensional vector, g is a continuous scalar function, and f is a continuous n-dimensional vector function determined by the dimension of the vector state variable w. Subsequently, a voltage-controlled memtransistor is defined[17] as:

$$\frac{dw}{dt} = f(w, V_G, V_D)$$

$$I_D = g(w, V_G, V_D) \quad (6)$$

$$I_G = h(w, V_G, V_D)$$

where h is a scalar function. The dependence of functions f, g and h on time t is determined from the voltage sweep rates. In first-order memristors, in which w is a scalar, w is usually defined by the physical region of devices with higher concentration of charged dopants or vacancies. Similarly, floating-gate metal-oxide-semiconductor FETs[16] and nanoionic dielectric gated FETs[15] have also been treated as first-order memtransistors, where f is a scalable function[17].

In the present MoS_2 memtransistors, EFM and cryogenic-transport measurements show strong evidence of a dynamic variation in Schottky barrier height. Indeed, Schottky barrier modulation is the underlying mechanism of interface-based switching in transition-metal-oxide and ferroelectric memristors, where local segregation of mobile dopants varies the tunnelling barrier dynamically[42–47]. Tuning the Schottky barrier by ion implantation (that is, by changing the local density of dopants) is central to contact engineering in FETs[26,27,38,39]. Therefore, we hypothesize that the local redistribution of dopants in polycrystalline MoS_2, facilitated by grain boundaries, is the main cause of barrier modulation. Varying the density of dopants (n_1) in the region w from the contact

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

edge (Extended Data Fig. 9a) changes the maximum electric field ξ_m (and the effective Schottky barrier height $\Delta\Phi_{image1}$) by image charge lowering[27] as:

$$|\xi_m| = \frac{e}{\varepsilon_s}(n_1 w - n(d-w)) \approx \frac{e n_1 w}{\varepsilon_s} \tag{7}$$

$$\Delta\Phi_{image1} \approx \frac{e}{\varepsilon_s}\sqrt{\frac{n_1 w}{4\pi}} \tag{8}$$

where d is the depletion width, n_1 is the net excess density near the contact, n is the dopant density in the channel away from the contacts and $\Delta n = n_1 - n$, as defined in Extended Data Fig. 7a. In the case of a 2D semiconductor, the metal image plane is cut by half, and thus equation (8) for a bulk semiconductor changes only by a factor of 2, which is neglected here because n_1 is a fitting parameter. However, the lowering of the image charge also depends on the drain bias V_D and the gate bias V_G and cannot be neglected at large biases. As shown previously[27,40,41], the easiest way to incorporate a high-bias scenario is through the geometrical factors A and B, which results in:

$$\Delta\Phi_{image2} = \sqrt{\frac{e\xi_m}{4\pi\varepsilon_s}} = \sqrt{\frac{e}{4\pi\varepsilon_s}}\left(\sqrt{\frac{B(V_G - V_{th})}{t_{ox}}} + \sqrt[4]{\frac{2en(\phi_{b0} + AV_D)}{\varepsilon_s}}\right) \tag{9}$$

where t_{ox} is the thickness of the oxide dielectric. Furthermore, at high biases, tunnelling cannot be neglected and its main features can be described by the following expression[40]:

$$\Delta\Phi_{tunnel} = \frac{1}{e}\left[\frac{3e\hbar(\ln 2)}{4\sqrt{2m^*}}\right]^{2/3}\left[\frac{B(V_G - V_{th})}{t_{ox}}\right]^{2/3} \tag{10}$$

where m^* is the effective mass and $\hbar$ is the Planck constant. The net change in effective barrier height ($\Delta\Phi$) can be determined by adding equations (8), (9) and (10) to obtain

$$\Delta\Phi = \Delta\Phi_{image1} + \Delta\Phi_{image2} + \Delta\Phi_{tunnel} \tag{11}$$

$$\begin{aligned}\Delta\Phi = {} & \frac{e}{\varepsilon_s}\sqrt{\frac{n_1 w}{4\pi}} + \sqrt{\frac{e}{4\pi\varepsilon_s}}\left(\sqrt{\frac{B(V_G - V_{th})}{t_{ox}}} + \sqrt[4]{\frac{2en(\phi_{b0} + AV_D)}{\varepsilon_s}}\right) \\ & + \frac{1}{e}\left(\frac{3e\hbar(\ln 2)}{4\sqrt{2m^*}}\right)^{2/3}\left(\frac{B(V_G - V_{th})}{t_{ox}}\right)^{2/3}\end{aligned} \tag{12}$$

A dynamically varying barrier changes the transistor's $I_D - V_D$ curve to yield a pinched hysteresis loop. Because the memristive switching ratio is largest in the sub-threshold regime (Fig. 2b), we take the following expression for the $I_D - V_D$ relationship of SB-FETs[48]:

$$I_D = D\exp\left[\frac{e(V_G - V_{th})}{c_r k_B T}\right]\left[1 - \exp\left(-\frac{eV_D}{k_B T}\right)\right] \tag{13}$$

where the first exponential term gives a straight line in the logarithmic–linear plot of $I_D - V_G$ (Fig. 2c) and the corresponding (sub-threshold) slope is given by $2.3\frac{k_B T}{e}c_r$. The second exponential term of equation (13) is responsible for the large asymmetry observed in the high-bias curve (Fig. 2a). The second term becomes negligible for $V_D > 3k_B T$, and thus the forward-bias current saturates quickly with increasing V_D (Fig. 2b). Because the variable-temperature conductivity fits well to the thermionic emission model (Fig. 3c) in the sub-threshold regime ($V_G < V_{th}$), we obtain the equation of the voltage-controlled memtransistor by multiplying equation (13) with the barrier height term from the thermionic equation (4):

$$I_D = D\exp\left[\frac{e(V_G - V_{th})}{c_r k_B T}\right]\left[1 - \exp\left(-\frac{eV_D}{k_B T}\right)\right]\exp\left(\frac{\Phi_b(w)}{k_B T}\right) \tag{14}$$

where $\Phi_b = \phi_{b0} - \Delta\Phi$, as shown schematically in Extended Data Fig. 7a, and ϕ_{b0} and Φ_b are the effective barrier heights before and after image lowering by $\Delta\Phi$, respectively. For $V_D > 0$ V, the Schottky diode at the source is reverse biased and acts as a bottleneck for the device current, whereas, for $V_D < 0$ V, the Schottky diode at the drain is reverse biased. Therefore, dopant redistribution at the two different contacts dominates in the positive and negative bias regimes, and the device can be visualized as two memristors connected by a transistor (Extended Data Fig. 8e). Local electrostatics at the two contacts is inherently asymmetric at large positive and large negative V_D values because $V_G - V_S < 0$ V at the source for $V_D > 0$ V (we note that $V_S = 0$ V), while $V_G - V_D > 0$ V at the drain for $V_D < 0$ V in the most relevant regime of $V_G < 0$ V and $|V_D| > |V_G|$. In other words, the Fermi level of MoS_2 is located deeper into the bandgap at the source in the forward bias than at the drain in the reverse bias. Because both V_D and V_G define the Schottky barrier modulation, we need two different sets of equations for the source and drain in the forward and reverse biases. Neglecting the gate leakage current I_G (Extended Data Fig. 3a), we modify equation (6) to get equations (15), where the only differences between functions f and g and between h and j are the fitting parameters. w_d and w_s are state variables at the drain and source contacts, respectively.

$$\begin{aligned}\frac{\partial w_d}{\partial t} &= f(w_d, V_D, V_G, t) \\ \frac{\partial w_s}{\partial t} &= g(w_s, V_D, V_G, t) \\ I_D &= \begin{cases} h(w_s, V_D, V_G, t) \text{ for } V_D \geq 0 \\ j(w_d, V_D, V_G, t) \text{ for } V_D < 0 \end{cases} \\ I_G &= 0\end{aligned} \tag{15}$$

Dynamic redistribution of dopants under bias is usually governed by drift motion. Consequently, physical models of memristors often require a nonlinear drift with the field to keep the state variable w within the boundary conditions. Dopant drift is greatly suppressed as it approaches the dimension of the device on either side, thereby avoiding irreversible hard switching or breakdown. The nonlinearity of dopant drift is usually modelled by a window function $F(w)$, and thus we obtain[1]:

$$\frac{\partial w}{\partial t} = \frac{\mu R_{on} I(t)}{d^2} F(w) \tag{16}$$

where d is the range of dopant drift, μ is the dopant mobility, $I(t)$ is the current as defined in equation (5) and R_{on} is the resistance of the device when w spans the whole distance d. The window function[49]:

$$F(w) = 1 - [(w - 0.5)^2 + 0.75]^p \tag{17}$$

has reproduced bipolar resistive switching in a wide range of metal-oxide memristors. It should be noted that larger p values imply larger nonlinearity. Thus, by taking $n_1 \approx \Delta n$ for large modulation in dopant density, we combine equations (12), (14), (15) and (17) to get the following set of equations for $V_D > 0$ V:

$$\begin{aligned}I_D = {} & D\exp\left[\frac{e(V_G - V_{th})}{c_r k_B T}\right]\left[1 - \exp\left(-\frac{e|V_D|}{c_{vd} k_B T}\right)\right] \\ & \times \exp\left(\frac{\phi_{b0} - \frac{e}{\varepsilon_s}\sqrt{\frac{w_s \Delta n}{4\pi}} + \sqrt{\frac{e}{4\pi\varepsilon_s}}\sqrt[4]{\frac{2en(\phi_{b0} + A|V_D|)}{\varepsilon_s}}}{k_B T}\right)\end{aligned} \tag{18}$$

$$\frac{\partial w_s}{\partial t} = EI_D\{1 - [(w_s - 0.5)^2 + 0.75]^p\} \tag{19}$$

$$V_D = \begin{cases} 10t & \text{for } t = 0 \text{ s to } 8 \text{ s} \\ 80 - 10(t - 8) & \text{for } t = 8 \text{ s to } 16 \text{ s} \end{cases} \tag{20}$$

Equation (20) describes the drain bias sweep, and A, D, E and Δn are fitting parameters. The parameter $p = 4$ provides sufficient non-linearity. The geometrical factor c_{vd} has a role similar to that of c_r and limits the exponentially growing current at high negative biases. Here, we first focus on fitting the $I_D - V_D$ curve for a fixed V_G; therefore, the terms involving the geometrical factor B in equation (12) are absorbed in the fitting parameter D. Similarity, the following set of equations describes the behaviour of the memtransistor for $V_D < 0$ V:

$$\begin{aligned}I_D = {} & D'\exp\left[\frac{e(V_G - V_{th})}{c_r k_B T}\right]\left[1 - \exp\left(\frac{e|V_D|}{c_{vd} k_B T}\right)\right] \\ & \times \exp\left(\frac{\phi'_{b0} - \frac{e}{\varepsilon_s}\sqrt{\frac{w_d \Delta n}{4\pi}} + \sqrt{\frac{e}{4\pi\varepsilon_s}}\sqrt[4]{\frac{2en(\phi_{b0} + A'|V_D|)}{\varepsilon_s}}}{k_B T}\right)\end{aligned} \tag{21}$$

$$\frac{\partial w_d}{\partial t} = E'I_D\{1 - [(w_d - 0.5)^2 + 0.75]^p\} \tag{22}$$

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

$$V_D = \begin{cases} -10(t-16) & \text{for } t = 16 \text{ s to } 24 \text{ s} \\ -80 + 10(t-24) & \text{for } t = 24 \text{ s to } 32 \text{ s} \end{cases} \quad (23)$$

where a different set of fitting parameters, A', D' and E', is needed owing to asymmetric electrostatics and dopant kinetics.

The experimental memtransistor characteristics are fitted well with this model, as shown in Fig. 2a. Extended Data Fig. 7b and c shows the dynamic modulation of w_d and w_s with V_D. The experimental values $c_r = 83.3$ and $V_{th} = 20$ V were used in addition to the constants $\phi_{b0} = 385$ meV and $\varepsilon_s = 4$ for monolayer[50] MoS_2. The fitting parameters in Fig. 2a are $A = 10^{-5}$, $D = 5.5 \times 10^{-9}$ and $E = 7 \times 10^{-4}$. We note that although we treat w and Δn independently in the mathematical formalism, we cannot measure them independently. Thus, we treat the product $w\Delta n$ as an internal state variable. For the fitting in Fig. 2a, we get $w_s\Delta n < 10^{10}\,\mathrm{cm}^{-2}$ and $w_d\Delta n < 3.5 \times 10^9\,\mathrm{cm}^{-2}$ in order to get Schottky barrier heights covering the entire range of the experimental values (80–125 meV) shown in Fig. 3c (in the model, Δn has units of cm^{-3}). The fitting in Fig. 2a results in effective Schottky barrier heights ranging from 20 meV to 280 meV in the LRS and the HRS, respectively. Thus, overestimation of the simulated values in the HRS and underestimation in the LRS could result from non-idealities in the transistor transport model at high biases. Assuming that w is of the order of a few nanometres, we get maximum excess doping of $\Delta n \approx 10^{10}\,\mathrm{cm}^{-2}$, which is within the range of previous measurements of doping from MoS_2 defects[51].

The smaller value of $w_d\Delta n$ compared to $w_s\Delta n$ results from the smaller fields at the drain for $V_D < 0$ V than at the source for $V_D > 0$ V, which is consistent with the smaller hysteretic loop in the reverse bias. We note that although the net device current (I_D) is determined by w_s for $V_D > 0$ V and by w_d for $V_D < 0$ V, for the device to exhibit reversible behaviour, w_d (w_s) must return to its initial value when $V_D > 0$ V ($V_D < 0$ V). This is achieved by setting appropriate fitting parameters D, E (D', E'). In real devices, a large electric field near forward-biased Schottky diodes and redistribution of dopants to the initial configuration are also expected. Finally, by considering the first exponential term in equations (18) and (21), which contains V_G, we simulate the gate tunability of memtransistor characteristics that is consistent with the observed experimental behaviour (Extended Data Fig. 7d, e). Deviations between simulations and experiments become larger at larger V_G values owing to the decreased validity of the SB-FET model beyond the sub-threshold regime.

The switching mechanism. The switching mechanism of MoS_2 memtransistors can be understood by assuming two memristors at the source (S-memristor) and the drain (D-memristor) that are connected by a FET (Extended Data Fig. 8e). We note that gate-tunable resistive switching cannot be obtained by simply connecting two conventional memristors with a transistor. The Schottky barrier height modulation is closely connected to the operation of the SB-FET in the MoS_2 memtransistor. Thus, this distinction between memristor and transistor only serves to specify the switching mechanism. MoS_2 memtransistors with residue-free interfaces between the metal and MoS_2 (Fig. 2) undergo HRS-to-LRS switching at positive bias and LRS-to-HRS switching in negative bias and are called LRS–HRS memtransistors here. MoS_2 memtransistors with an approximately 1.5-nm-thick polymer tunnel barrier between the metal and MoS_2 (Extended Data Fig. 8a–d) undergo similar HRS-to-LRS switching at both positive and negative biases and are called LRS–LRS memtransistors. The main difference between the two types is that the LRS–HRS memtransistor retains its state while crossing the 0 V level, whereas LRS–LRS memtransistors undergo switching while crossing 0 V (Figs 2b, 3d). Both of these contrasting behaviours can be explained by the resistance table in Extended Data Fig. 8e. HD and LD are the resistance values of the HRS and LRS of the D-memristor, whereas HS and LS are the resistance values of the HRS and LRS of the S-memristor, respectively. We have HD >> LD and HS >> LS for both kinds of memtransistors, and their relative amplitudes determine the switching ratio. LD′ and LS′ are intermediate states with resistance values that differ vastly for LRS–HRS and LRS–LRS memtransistors.

For LRS–HRS memtransistors, switching events A and B dominate, while switching event C is negligible (Extended Data Fig. 8e). Switching A occurs at the bottleneck contact (that is, the source for forward bias and the drain for reverse bias), while switching B occurs at the other contact in order to restore the dopant distribution at the end of a full sweep cycle. This condition is necessitated by bipolar resistive switching, and its physical origin is well explained by the memtransistor model. Thus, intermediate states LD′ and LS′ have resistance values close to LD and LS, respectively.

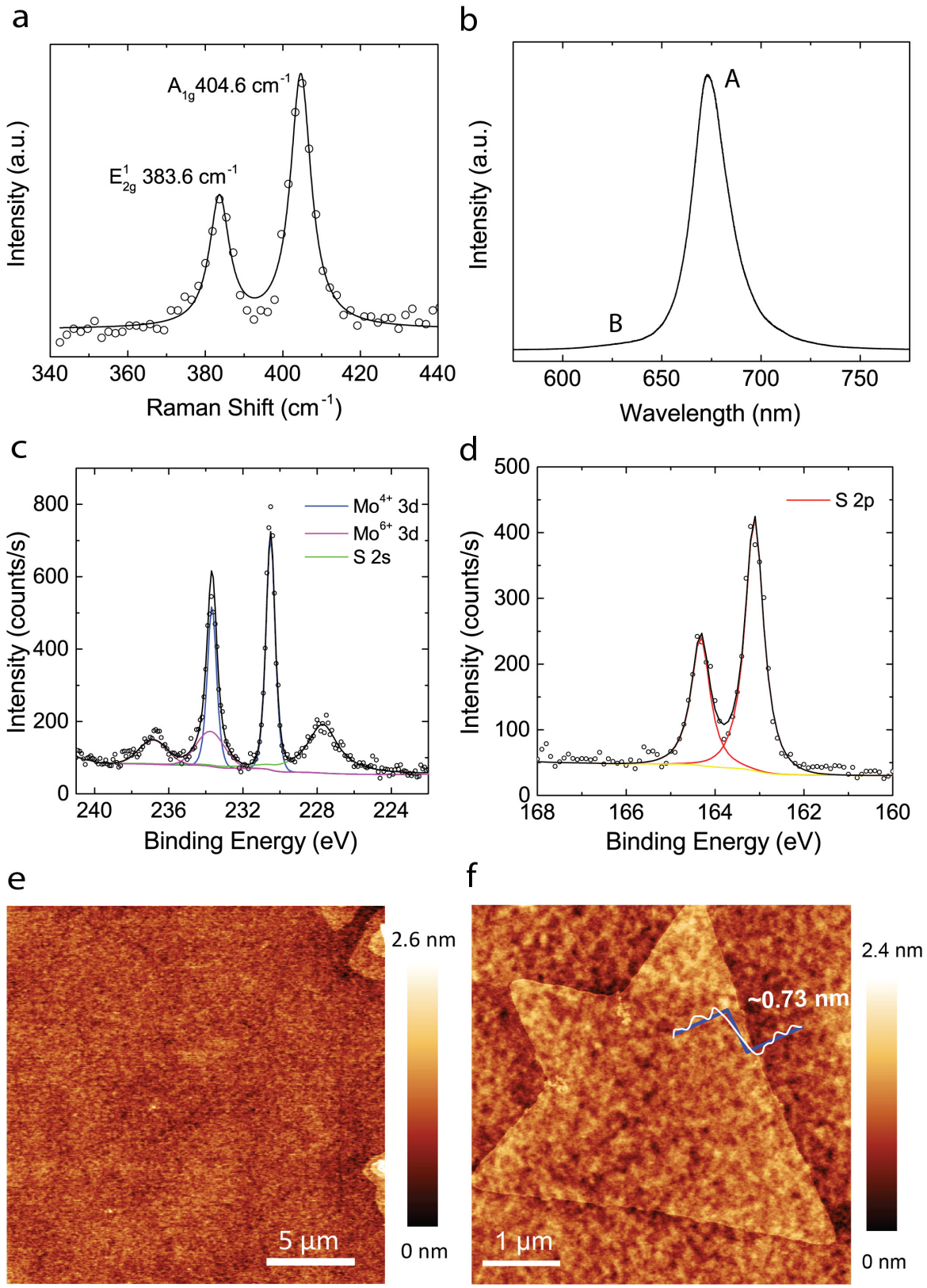
On the other hand, for LRS–LRS memtransistors, all switching events A, B and C appear. The resist layer acts at a tunnel barrier that minimizes pinning of the Fermi level of MoS_2. Similar minimization of the Fermi level has been shown in MoS_2 by employing ultrathin tunnel barriers, such as 2 nm MgO (ref. 35), 1 nm TiO_2 (ref. 52), 1.5 nm Ta_2O_5 (ref. 53) and 0.6 nm hexagonal boron nitride[54], resulting in reduction of the effective Schottky barrier height from 100 meV to 23 meV, from 121 meV to 27 meV, from 95 meV to 29 meV and from 159 meV to 31 meV, respectively. Introduction of the tunnel barrier increases the thermionic emission contribution to the total current considerably, outweighing the additional tunnel resistance. Here, a tunnel barrier of about 1.5-nm-thick polymer resist makes the resistances of reverse-biased Schottky diodes at the source ($V_D > 0$ V) and drain ($V_D < 0$ V) more symmetric, resulting in more symmetric I_D–V_D compared to LRS–HRS memtransistors (see Figs 3d, 2b). This switching involves a dynamic negative differential resistance feature for $V_D < 0$ V common in memristors (Fig. 3d)[5]. This phenomenon increases the resistance of the intermediate states LD′ and LS′ so that LD′ >> LD and LS′ >> LS. In other words, switching B in Extended Data Fig. 8e cannot be neglected, which explains the full switching cycles of LRS–LRS memtransistors. Compared to LRS–HRS memtransistors, LRS–LRS memtransistors show higher conductivity at the same V_D bias and switching at smaller V_G bias. This behaviour can be explained by the smaller contact resistance from the lack of Fermi level pinning or increased doping from the residue layer.

This analysis of the LRS–HRS memtransistor also sheds light on the operating mechanism of heterosynaptic multi-terminal devices (Fig. 4b). Conductance changes occur between the side electrodes 1–4 upon the application of high-bias pulses between the main electrodes (5, 6) owing to modulation of the Schottky barrier near the side electrodes. The switching ratios were observed to increase with the overlapping areas of the side electrode with the MoS_2. Thus, floating electrodes pin the Fermi level of MoS_2, and the energy level of MoS_2 under the side electrodes is lower than that of the MoS_2 region outside the side electrodes (Extended Data Fig. 10g, h), resulting in additional band-bending in the channel. Dopants are expected to be redistributed near the floating electrodes in the same manner as LRS–HRS memtransistors during sweeps 2 and 4 (Fig. 2a).

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

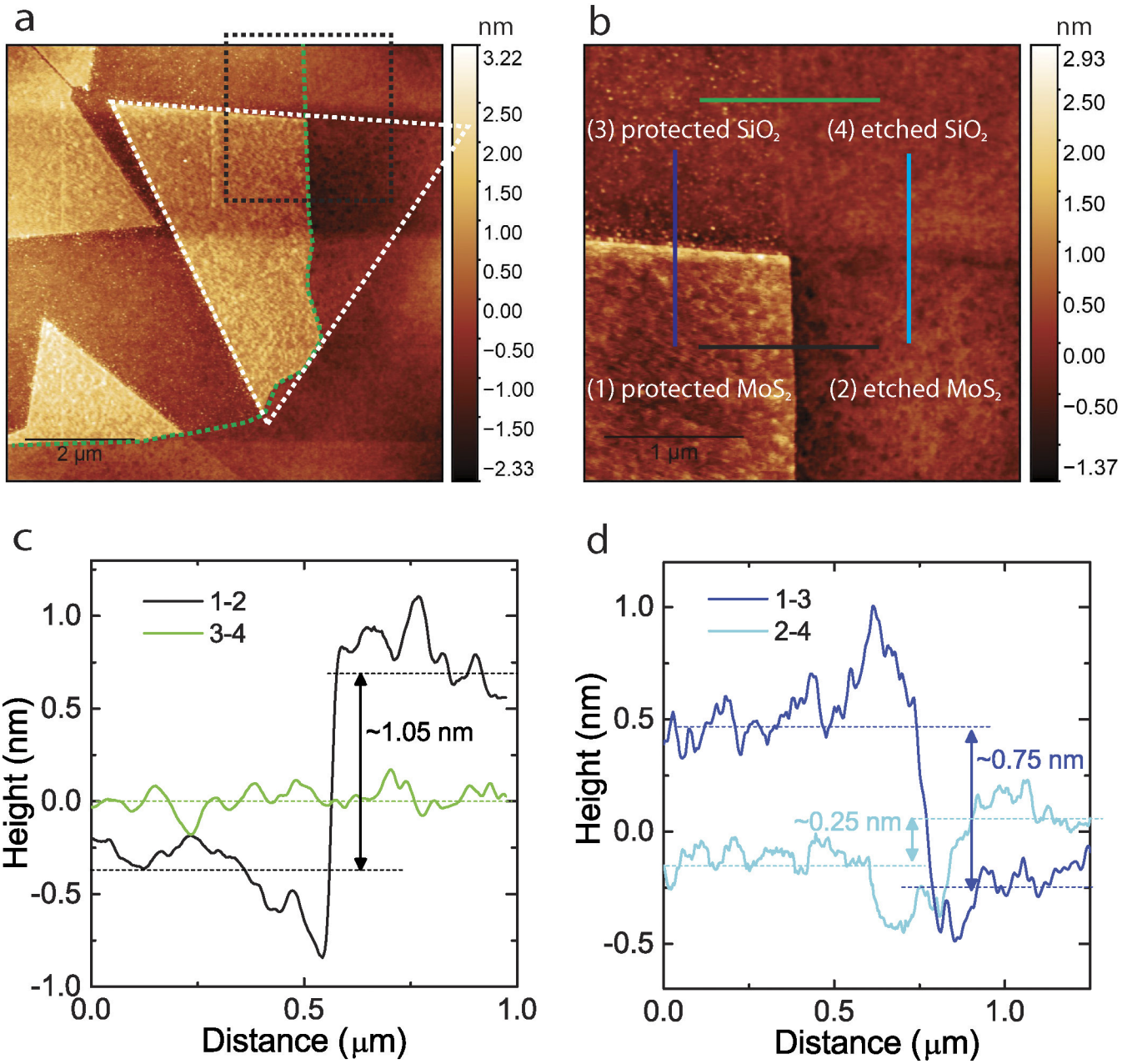
33. Shastry, T. A. *et al.* Mutual photoluminescence quenching and photovoltaic effect in large-area single-layer MoS_2–polymer heterojunctions. *ACS Nano* **10,** 10573–10579 (2016).
34. Lee, C. *et al.* Anomalous lattice vibrations of single- and few-layer MoS_2. *ACS Nano* **4,** 2695–2700 (2010).
35. Chen, J.-R. *et al.* Control of Schottky barriers in single layer MoS_2 transistors with ferromagnetic contacts. *Nano Lett.* **13,** 3106–3110 (2013).
36. Chua, L. O. & Sung Mo, K. Memristive devices and systems. *Proc. IEEE* **64,** 209–223 (1976).
37. Chua, L. Memristor – the missing circuit element. *IEEE Trans. Circuit Theory* **18,** 507–519 (1971).
38. Shannon, J. M. Reducing the effective height of a Schottky barrier using low-energy ion implantation. *Appl. Phys. Lett.* **24,** 369–371 (1974).
39. Shannon, J. M. Increasing the effective height of a Schottky barrier using low-energy ion implantation. *Appl. Phys. Lett.* **25,** 75–77 (1974).
40. Kedzierski, J. *et al. Int. Electron Devices Meeting 2000* 57–60 http://ieeexplore.ieee.org/document/904258/ (IEEE, 2000).
41. Calvet, L. E. *Electrical transport in Schottky barrier MOSFETs.* https://www.eng.yale.edu/reedlab/publications/Calvet%20Thesis.pdf PhD thesis, Yale Univ. (2001).
42. Sawa, A. Resistive switching in transition metal oxides. *Mater. Today* **11,** 28–36 (2008).
43. Kubicek, M., Schmitt, R., Messerschmitt, F. & Rupp, J. L. M. Uncovering two competing switching mechanisms for epitaxial and ultrathin strontium titanate-based resistive switching bits. *ACS Nano* **9,** 10737–10748 (2015).
44. Lee, H.-S. *et al.* Ferroelectric tunnel junction for dense cross-point arrays. *ACS Appl. Mater. Interfaces* **7,** 22348–22354 (2015).
45. Kim, D. J. *et al.* Ferroelectric tunnel memristor. *Nano Lett.* **12,** 5697–5702 (2012).
46. Nakamura, M. *et al.* Interface band profiles of Mott-insulator/Nb:$SrTiO_3$ heterojunctions as investigated by optical spectroscopy. *Phys. Rev. B* **82,** 201101 (2010).
47. Sawa, A., Fujii, T., Kawasaki, M. & Tokura, Y. Hysteretic current–voltage characteristics and resistance switching at a rectifying Ti/$Pr_{0.7}Ca_{0.3}MnO_3$ interface. *Appl. Phys. Lett.* **85,** 4073–4075 (2004).
48. Streetman, B. G. & Banerjee, S. K. *Solid State Electronic Devices* (Prentice Hall, 1995).
49. Prodromakis, T., Peh, B. P., Papavassiliou, C. & Toumazou, C. A versatile memristor model with nonlinear dopant kinetics. *IEEE Trans. Electron Dev.* **58,** 3099–3105 (2011).
50. Liu, W. *et al.* Impact of contact on the operation and performance of back-gated monolayer MoS_2 field-effect-transistors. *ACS Nano* **9,** 7904–7912 (2015).
51. Kim, I. S. *et al.* Influence of stoichiometry on the optical and electrical properties of chemical vapor deposition derived MoS_2. *ACS Nano* **8,** 10551–10558 (2014).
52. Dankert, A., Langouche, L., Kamalakar, M. V. & Dash, S. P. High-performance molybdenum disulfide field-effect transistors with spin tunnel contacts. *ACS Nano* **8,** 476–482 (2014).
53. Lee, S., Tang, A., Aloni, S. & Wong, H. S. P. Statistical study on the Schottky barrier reduction of tunneling contacts to CVD synthesized MoS_2. *Nano Lett.* **16,** 276–281 (2016).
54. Wang, J. *et al.* High mobility MoS_2 transistor with low Schottky barrier contact by using atomic thick h-BN as a tunneling layer. *Adv. Mater.* **28,** 8302–8308 (2016).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



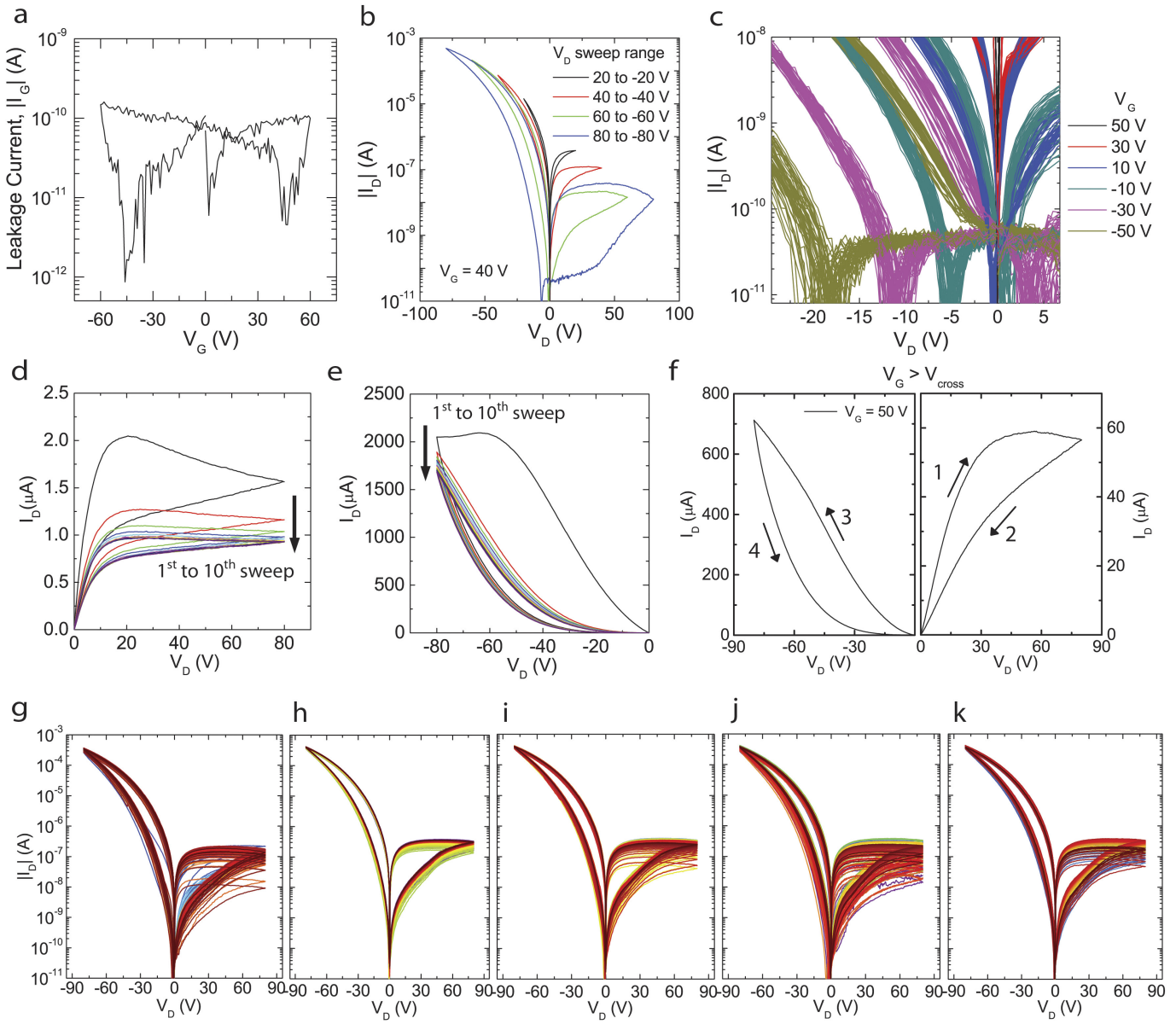
Extended Data Figure 1 | Material characterization of the MoS_2 film. **a**, Raman spectrum of CVD-grown polycrystalline monolayer MoS_2, measured using an excitation wavelength of 532 nm. The Lorentzian peak fits correspond to the E^1_{2g} and A_{1g} modes. **b**, Photoluminescence spectrum of MoS_2 collected with the same microscope. **c**, **d**, XPS spectra of MoS_2 on a SiO_2/Si substrate, showing the Mo 3*d*, S 2*s* and S 2*p* peaks. **e**, AFM topography image corresponding to the lateral force microscopy image of Fig. 1c. **f**, AFM topography image of the edge of a MoS_2 flake, showing a monolayer step height of about 0.73 nm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



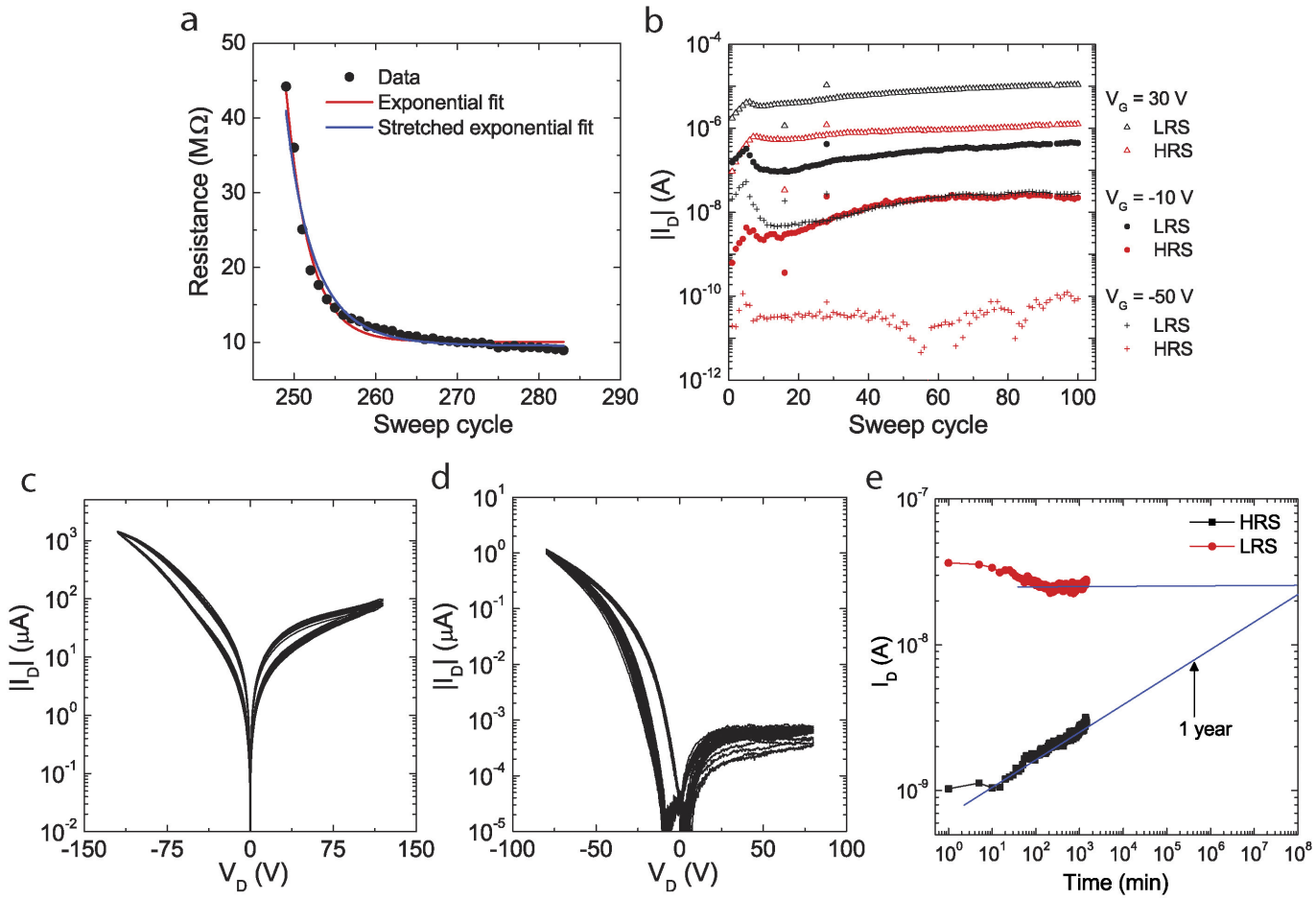
Extended Data Figure 2 | AFM analysis of a residue-free photolithography process. a, AFM topography image of MoS_2 crystals patterned by PMGI-assisted photolithography. The dashed green line shows the location of the edge of the patterned photoresist in the left region and the white dashed line shows the triangular MoS_2 crystal domain before reactive ion etching. **b**, Magnified AFM topography image of the region defined by the black dashed line in **a**, showing chequered regions of protected (1) MoS_2, (2) etched MoS_2, (3) protected SiO_2 and (4) etched SiO_2. **c**, Height profiles taken along the two horizontal lines in **b**, showing minimal residue left on the protected SiO_2 region. **d**, Height profiles taken along the two vertical lines in **b**, showing minor etching of SiO_2 under the etched MoS_2 region (2). The noise in the height profiles is due to surface roughness and tip artefacts.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



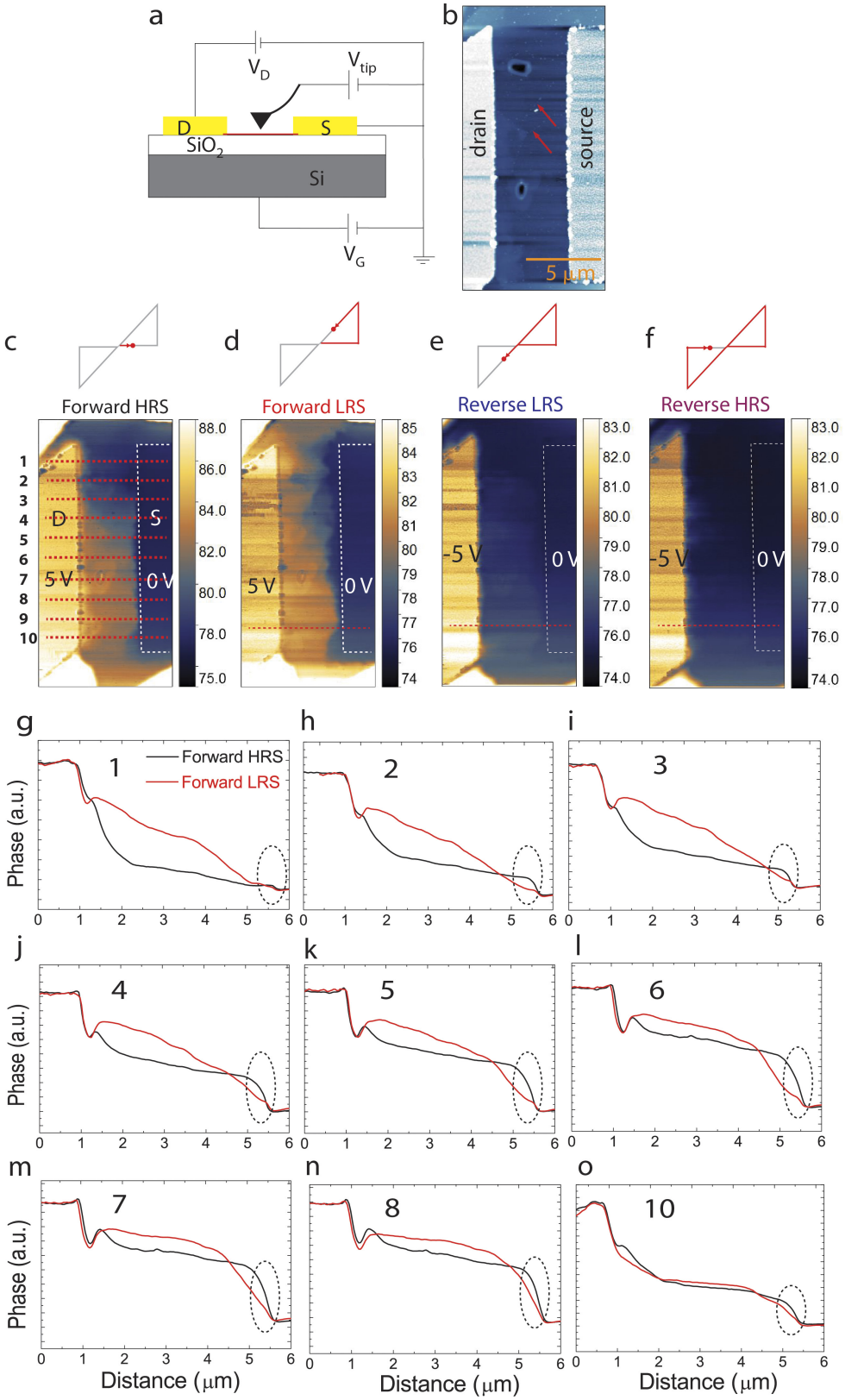
Extended Data Figure 3 | Extended electrical characteristics of the MoS_2 memtransistor. a, Leakage current I_G of the MoS_2 memtransistor of Fig. 2a as a function of V_G after a high-bias sweep from $V_D = 80$ V to $V_D = -80$ V. We note that the current level of 100 pA is close to the instrumentation noise floor. **b**, I_D–V_D curve of a memtransistor ($L = 15\,\mu m$, $W = 150\,\mu m$) for different V_D sweeps from |20| V to |80| V, showing increasing switching ratio with sweep range (switching ratio $>10^3$ for the range from 80 V to −80 V). **c**, Magnified view of 50 sweep cycles of the device from Fig. 2b, showing an insulating state in a range of negative V_D values that is dependent on V_G and non-zero crossing, suggesting memcapacitance from contacts. **d**, I_D–V_D curve of a MoS_2 memtransistor during ten consecutive unipolar positive-bias sweeps from $V_D = 0$ V to 80 V. **e**, I_D–V_D curve of the same MoS_2 memtransistor during ten consecutive unipolar negative-bias sweeps from $V_D = 0$ V to –80 V. **f**, Switching from the LRS to the HRS for the MoS_2 memtransistor of Fig. 2b in the forward bias for $V_G > V_{cross}$, where $V_{cross} \approx 35$ V. I_D–V_D curve of the device of Fig. 2d during 475 voltage sweeps: **g**, sweeps 1–100; **h**, sweeps 100–200; **i**, sweeps 200–300; **j**, sweeps 300–400; **k**, sweeps 400–475.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



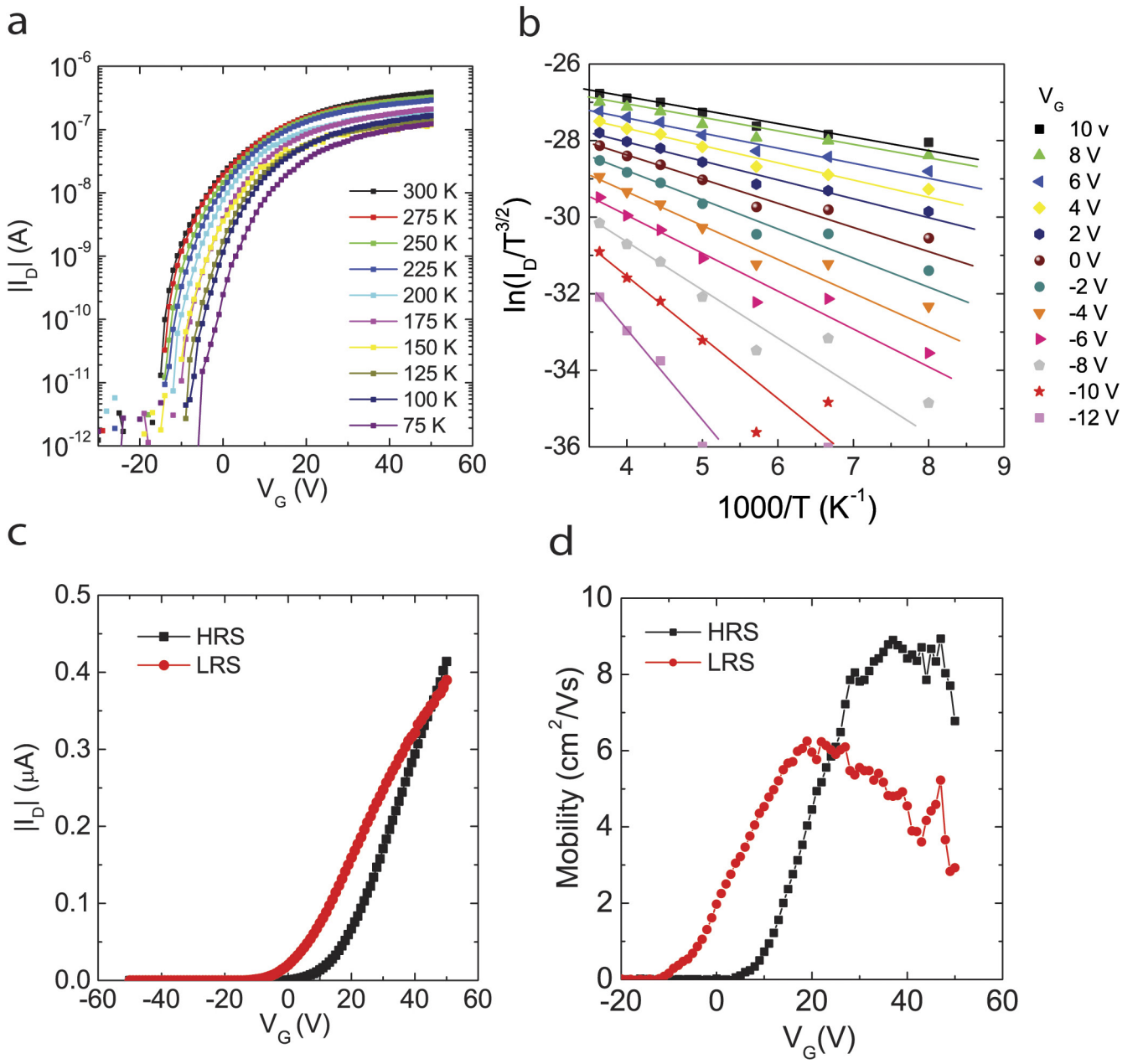
Extended Data Figure 4 | Current endurance characteristics. **a**, Exponential and stretched exponential fits to a typical subset of endurance points from Fig. 2d. The stretched exponential function is defined as $I_D \approx A - Be^{C(n_0 - n)^\gamma}$, where A, B, C and n_0 are constants and $\gamma \approx 0.8$. Both the exponential and stretched exponential fits show $R^2 \approx 0.97$, but the stretched exponential shows a better fit at the tail end of the curve. **b**, Endurance characteristics of a memtransistor, showing only one exponential decay in reverse bias ($V_D = -10$ V). **c**, I_D–V_D curve ($V_G = 0$ V) of a device with $L = 20\,\mu$m and $W = 150\,\mu$m, showing a negligible memristive loop (ten sweep cycles) for an unoptimized geometry. **d**, I_D–V_D curve ($V_G = 60$ V) of a device with $L = 10\,\mu$m and $W = 5\,\mu$m, showing a negligible memristive loop (19 sweep cycles) for an unoptimized geometry. **e**, HRS and LRS retention characteristics from Fig. 2e plotted and extrapolated in a doubly logarithmic scale. The relaxation of the two states is faster than conventional filament-based memristors, such as TiO_2.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



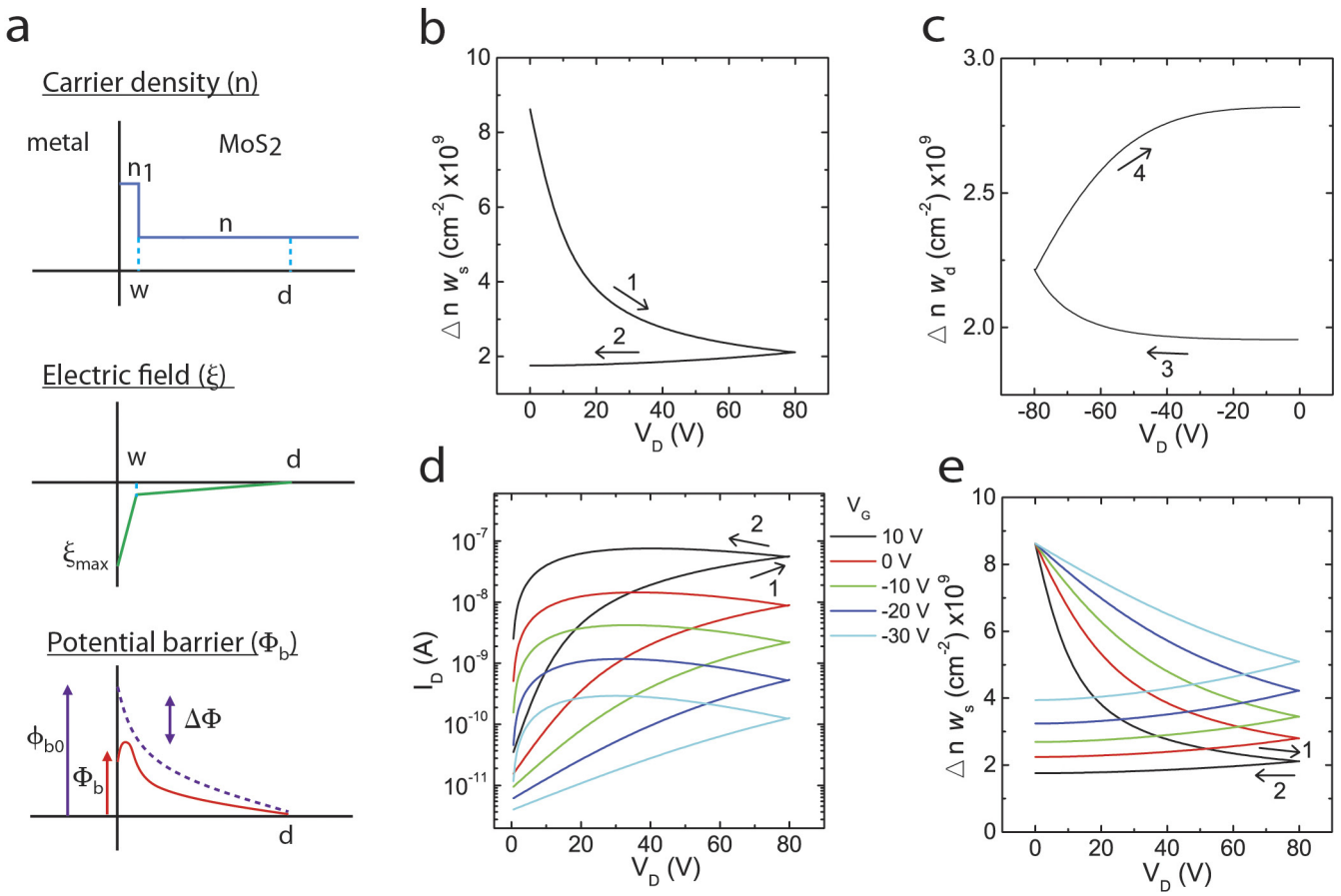
Extended Data Figure 5 | *In situ* EFM of a MoS_2 memtransistor. **a**, Schematic of the *in situ* EFM measurements of MoS_2 memtransistors. **b**, AFM topography image of the device from Fig. 3a, showing grain boundaries highlighted by red arrows. **c**, Reproduction of the EFM phase images of Fig. 3a in the forward HRS. **d**–**f**, EFM phase images in the forward LRS, reverse LRS and reverse HRS, which were used for the line profiles shown in Fig. 3b. **g**–**o**, EFM phase profiles along the red dashed lines 1–8 and 10 in **c** and **d**. The EFM phase profile along line 9 is shown in Fig. 3b. All profiles are averaged over 128 lines and are normalized with the EFM phase values at the drain and source.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



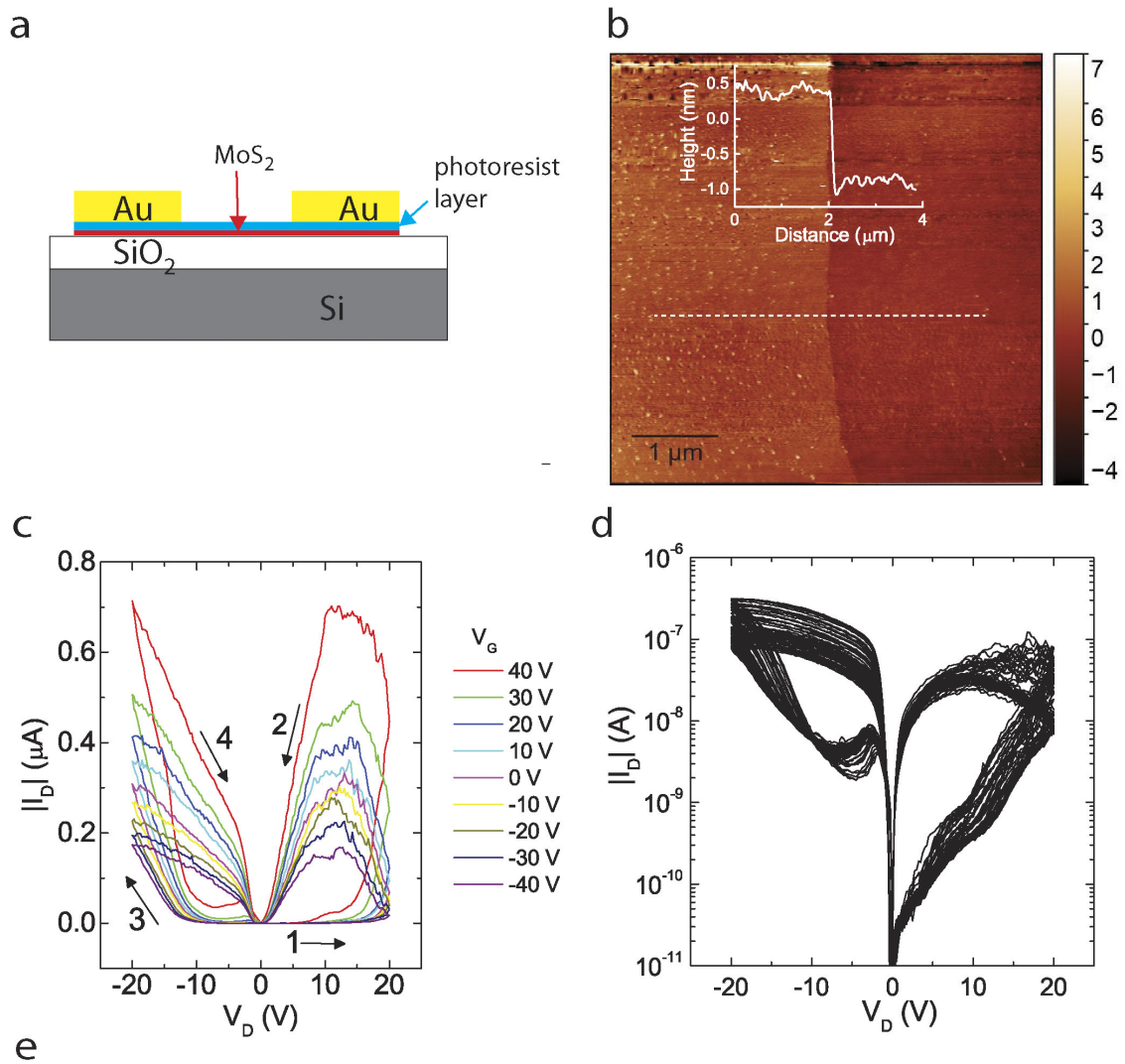
Extended Data Figure 6 | Low-temperature transport measurements of a MoS_2 memtransistor. a, I_D–V_G curve of the device shown in Fig. 3c (in the LRS) at $V_D = 0.1$ V for temperature varying from 300 K to 75 K at a step of 25 K. **b**, Plot of $\ln(I_D/T^{3/2})$ versus 1,000/T for different V_G values, which was used to extract the Schottky barrier height through the thermionic emission model. **c**, **d**, I_D–V_G and field-effect mobility versus the V_G value of the same device in the LRS and the HRS, respectively. The crossing curves in **c** show V_{th} shifts by 15 V between the HRS and LRS and $V_{cross} \approx 42$ V.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

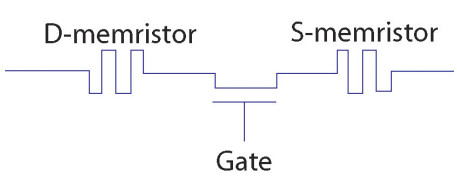


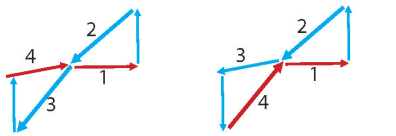
Extended Data Figure 7 | Memtransistor modelling. a, Schematic of the increased doping region near the contact, which results in a larger field and reduced metal–semiconductor Schottky barrier height. **b**, Simulated variation of $w_s\Delta n$ at the source contact for forward bias (sweeps 1 and 2 in Fig. 2a). **c**, Simulation variation of $w_d\Delta n$ at the drain contact for reverse bias (sweeps 3 and 4 in Fig. 2a). **d**, Simulated I_D–V_D curve of a MoS_2 memtransistor in the forward bias with different V_G values from 10 V to −30 V. **e**, Simulated variation in $w_s\Delta n$ for the same V_G values. The key between **d** and **e** applies to both plots.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



e

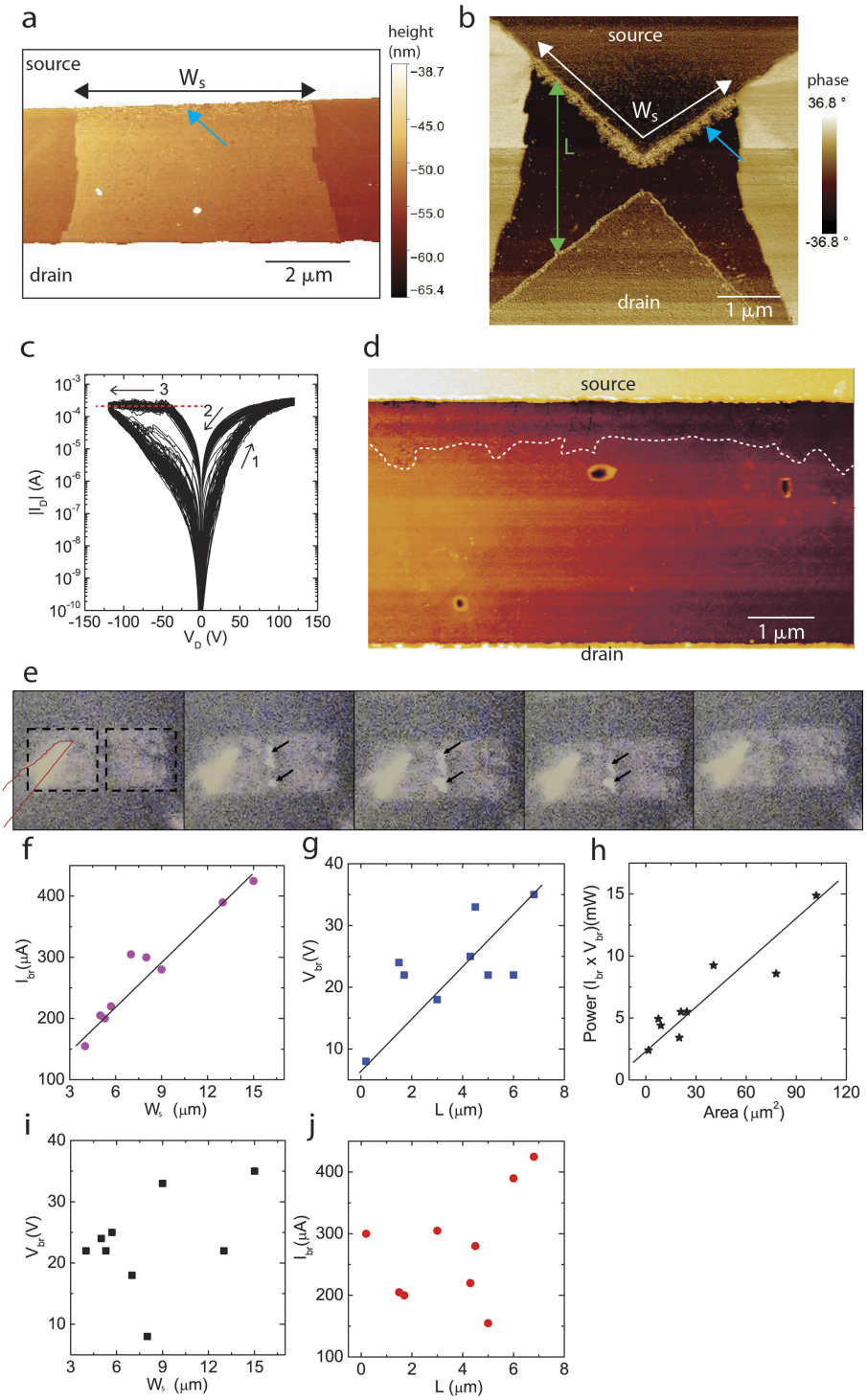




Sweep	Resistance D-memristor	Resistance S-memristor	Total Resistance	LRS-HRS Switching	LRS-LRS Switching
1	HD ↓	HS ↓ A	HD + HS	HRS	HRS
2	↓ LD'	LS ↓	LD' + LS	LRS	LRS
3	↓ LD	B ↓ LS	LD + LS	LRS	HRS
4	HD ↓	LS' ↓ C	HD + LS'	HRS	LRS
1	HD	HS	HD + HS	HRS	HRS

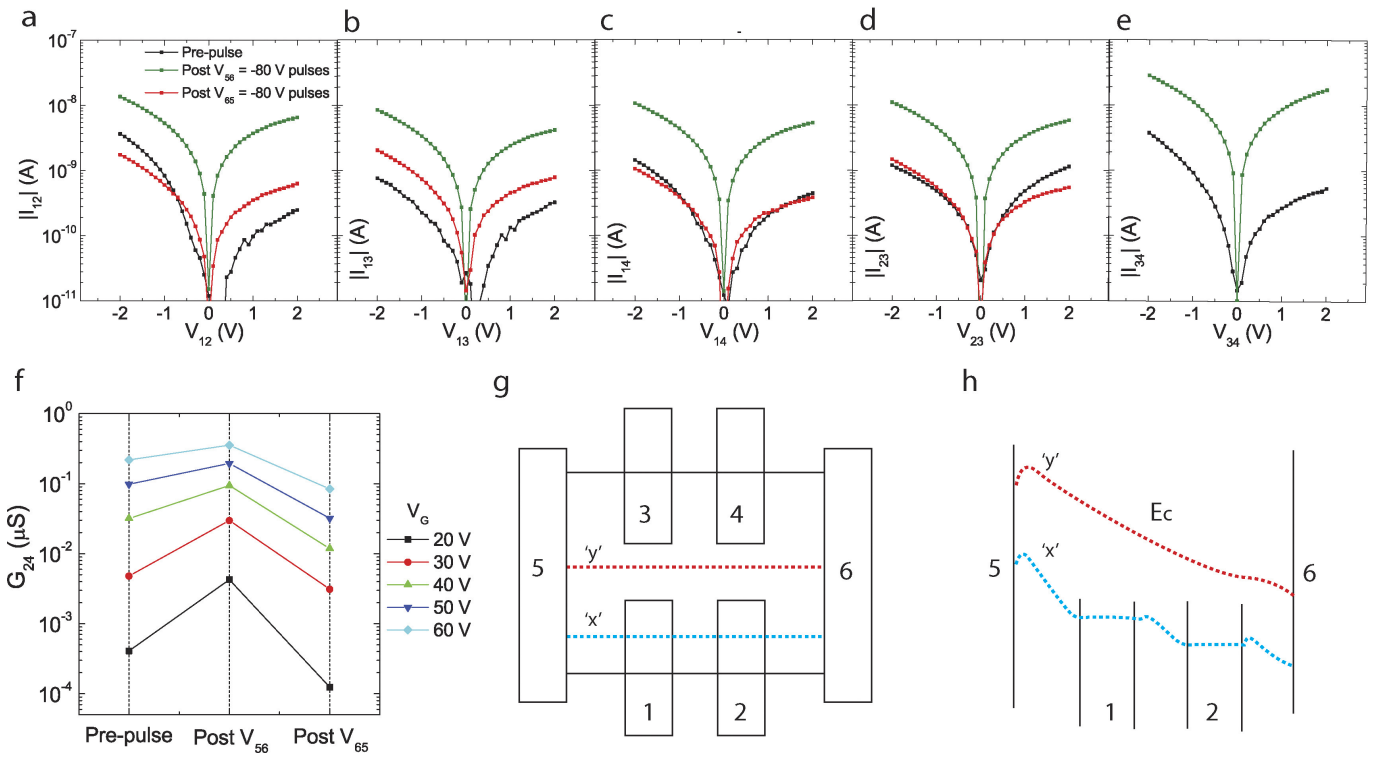
Extended Data Figure 8 | LRS–LRS MoS_2 memtransistor characteristics and mechanism. a, Schematic of an LRS–LRS memtransistor in which a thin photoresist layer acts as a tunnel barrier between the metal contacts and the MoS_2 film. **b**, AFM topography images, showing the step height of the remaining photoresist on a blank Si substrate after a fabrication process without using PMGI. The inset shows the height profile along the white dashed line, which reveals a thickness of about 1.5 nm. **c**, Gate-tunable I_D–V_D curves from Fig. 3d, shown in a linear scale. **d**, I_D–V_D curves of 50 sweep cycles of the LRS–LRS memristor of Fig. 3d. **e**, Table showing the resistive switching characteristics of LRS–HRS and LRS–LRS memtransistors at the source and drain contacts during the four stages of a full sweep cycle. The conditions of the relative resistance values that are necessary for the two different switching behaviours are listed in the top right corner. Three kinds of resistive switching events, A, B and C, are shown by coloured arrows (see Methods section 'The switching mechanism' for details).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 9 | Electromigration-induced degradation in control MoS_2 devices. a, AFM topography images (corresponding to the upper inset of Fig. 4a), showing electromigration-induced degradation in the material (cyan arrow) near the source electrode (top). The colour scale represents height difference. **b**, AFM phase image of a device with an hourglass-shaped channel (that is, varying channel length from 5 μm to 1 μm), showing dendritic features along the entire source electrode (top). Without Schottky contacts, we expect a thermal 'hot spot' with high local temperature in the region of the highest electric field (V_D/L) (that is, only in the narrowest region in the centre of the channel). Absence of such localized breakdown rules out Joule heating and favours electromigration near the source contact as the dominant phenomenon. The width of the source electrode edge (W_s) is shown by the white arrows. **c**, I_D–V_D curves (85 sweep cycles) of a degraded polycrystalline monolayer MoS_2 memtransistor at $V_G = 0$ V ($L = 5$ μm, $W = 100$ μm). **d**, AFM topography image of the device of **c**, showing the dendritic features above the white dashed line. **e**, A series of five successive snapshots (left to right) from a video captured by a black-and-white camera during sweep 3, as indicated by the dashed red line in **c**. The red outline and dashed black line in the first frame show the probe tip and electrode pads, respectively. The three middle frames show bright spots from light emission in the channel close to the source electrode (right), marked by black arrows. Light emission was observed during all 85 sweep cycles shown in **c**. **f**, Breakdown current I_{br} (defined in Fig. 4a), showing a linear correlation with W_s for nine single-flake control MoS_2 devices. **g**, Breakdown voltage V_{br} (defined in Fig. 4a), showing a linear correlation with L, which suggests that the potential decreases both across the channel and at the Schottky contacts. **h**, Power ($I_{br} \times V_{br}$), showing a linear correlation with the channel area ($L \times W$). **i**, V_{br}, showing no correlation with W_s. **j**, I_{br} also shows no correlation with L.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | Electrical characteristics of multi-terminal heterosynaptic device. a–e, Low-bias I_{ij}–V_{ij} curves of 5 permutations of the inner electrodes 1–4 of Fig. 4b at $V_G = 20$ V, where $ij = 12$, 13, 14, 23 and 34. I_{24}–V_{24} is shown in Fig. 4b. The key in **a** applies to all panels **a–e**. The black curves were measured before any pulsing. The green curves were measured after applying a −80-V pulse at V_{56} (5, drain; 6, source) four times at a voltage ramping rate of 10 V s^{-1}. The red curves were measured after applying a −80-V pulse at V_{65} (6, drain; 5, course) three times at the same ramping rate. The conductance returns to the pre-pulse state for all electrode combinations. The I_{34}–V_{34} curve could not be measured after the V_{65} pulse cycle. **f**, Change in the conductance between electrodes 2 and 4 (G_{24}), with V_{56} and V_{65} pulses for different V_G values showing gate tunability of heterosynaptic plasticity. **g**, **h**, Spatial profile of the MoS_2 conductance band minimum (E_c) along the two dashed lines in **g**, which pass through (*x*) and outside (*y*) the side electrodes.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature25765

Sterically controlled mechanochemistry under hydrostatic pressure

Hao Yan[1,2]*, Fan Yang[1,3]*, Ding Pan[4,5,6]*, Yu Lin[1], J. Nathan Hohman[7], Diego Solis-Ibarra[8], Fei Hua Li[1,2], Jeremy E. P. Dahl[1], Robert M. K. Carlson[1], Boryslav A. Tkachenko[9], Andrey A. Fokin[9,10], Peter R. Schreiner[9], Giulia Galli[11], Wendy L. Mao[1,3], Zhi-Xun Shen[1,12] & Nicholas A. Melosh[1,2]

Mechanical stimuli can modify the energy landscape of chemical reactions and enable reaction pathways, offering a synthetic strategy that complements conventional chemistry[1–3]. These mechanochemical mechanisms have been studied extensively in one-dimensional polymers under tensile stress[4–9] using ring-opening[10] and reorganization[11], polymer unzipping[6,12] and disulfide reduction[13,14] as model reactions. In these systems, the pulling force stretches chemical bonds, initiating the reaction. Additionally, it has been shown that forces orthogonal to the chemical bonds can alter the rate of bond dissociation[15]. However, these bond activation mechanisms have not been possible under isotropic, compressive stress (that is, hydrostatic pressure). Here we show that mechanochemistry through isotropic compression is possible by molecularly engineering structures that can translate macroscopic isotropic stress into molecular-level anisotropic strain. We engineer molecules with mechanically heterogeneous components—a compressible ('soft') mechanophore and incompressible ('hard') ligands. In these 'molecular anvils', isotropic stress leads to relative motions of the rigid ligands, anisotropically deforming the compressible mechanophore and activating bonds. Conversely, rigid ligands in steric contact impede relative motion, blocking reactivity. We combine experiments and computations to demonstrate hydrostatic-pressure-driven redox reactions in metal–organic chalcogenides that incorporate molecular elements that have heterogeneous compressibility[16–19], in which bending of bond angles or shearing of adjacent chains activates the metal–chalcogen bonds, leading to the formation of the elemental metal. These results reveal an unexplored reaction mechanism and suggest possible strategies for high-specificity mechanosynthesis.

We demonstrate the molecular anvil concept (Extended Data Fig. 1) in copper(I) *m*-carborane-9-thiolate (Cu-S-M9) crystals (where *m* denotes the meta positions of the carbon atoms in the carborane), synthesized via ligand-directed assembly[16]. The unit cell (Fig. 1a, Extended Data Fig. 2 and Supplementary Table 1), determined by single-crystal X-ray diffraction (XRD), is composed of two enantiomers. Each molecule (Fig. 1b) has a Cu_4S_4 core surrounded by M9 ligands. These ligands are anticipated to be much more structurally rigid, owing to their cage-like nature, than the Cu_4S_4 core. Importantly, the distances between M9 groups (7.4–9.2 Å) are substantially larger than the molecular distance in *m*-carborane crystals (6.9 Å, ref. 17), indicating that the M9 groups are not in van der Waals contact, thus allowing them to move without mutual steric hindrance under compression.

We first show that hydrostatic pressure drives a Cu(I) to Cu(0) reduction in Cu-S-M9. Transmission electron microscopy (TEM) images of a sample after compression to 12 GPa reveals crystalline nanoclusters (Fig. 1c and d) with lattice spacing of 2.08 Å, consistent with the [111] spacing of face-centred cubic copper. Energy dispersive X-ray spectroscopy (EDS, Fig. 1e) shows that the nanoclusters are composed exclusively of copper and are free of sulfur, clearly different from the pristine Cu-S-M9 (Extended Data Fig. 3). These results indicate that Cu(I) in Cu-S-M9 is reduced to Cu(0) under hydrostatic pressure. Note that the average size of the Cu nanoclusters, about 10 nm, is not determined by the size of the starting crystal (1–100 μm, inset to Fig. 1c), but rather by the nucleation density and diffusivity of atomic copper at high pressure. Here the sulfur serves as the reducing agent, forming an oxidized species (for example, disulfide[20]), as revealed by X-ray photoelectron spectroscopy (Extended Data Fig. 4) and supported by density functional theory (DFT) computations discussed later.

This pressure-driven mechanochemical reaction takes a different pathway from conventional thermochemistry of the same compound. Heating Cu-S-M9 to 400 °C under vacuum yields cuprous sulfide (Extended Data Fig. 5) instead of elemental copper. Literature data also indicate that pyrolysis of a chemically similar compound, copper(I) *tert*-butylthiolate, yields cuprous sulfide[21].

The reactivity of the Cu-S-M9 system is pressure-dependent. We first note that elemental copper was observed only after compression beyond 8 GPa (Extended Data Fig. 3). To understand this threshold behaviour, we used *in situ* XRD to track the structural change of the system and its connection with reactivity. The XRD patterns below 8 GPa show well defined peaks corresponding to Cu-S-M9, which shift smoothly to increasing 2θ (decreasing the lattice spacing d) as pressure increases (vertical dotted lines, Fig. 1f). The unit cell volumes can be fitted to a third-order Birch–Murnaghan equation of state[22] (solid red line, Fig. 1g), consistent with elastic compression. Releasing the pressure after compression within this range fully restores the ambient diffraction pattern (Extended Data Fig. 6), suggesting no irreversible structural change. However, compression beyond 8 GPa results in missing diffraction peaks, and at 12 GPa all diffraction peaks related to Cu-S-M9 disappear (purple, Fig. 1f), indicating a disordered phase. The diffraction pattern of Cu-S-M9 is not restored after compression to this range (green, Fig. 1f), indicating irreversible structural change. The pressure range in which the crystalline-disorder transition occurs, 8–12 GPa, agrees well with the pressure window for elemental copper formation observed by TEM.

To establish that this change is due to a pressure-driven reaction, we examined the valence change of copper as a function of pressure. *In situ* X-ray absorption spectroscopy (XAS) shows that the intensity of

[1]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. [2]Department of Materials Science and Engineering, Stanford University, Stanford, California 94305, USA. [3]Department of Geological Sciences, Stanford University, Stanford, California 94305, USA. [4]Department of Physics, Hong Kong University of Science and Technology, Hong Kong, China. [5]Department of Chemistry, Hong Kong University of Science and Technology, Hong Kong, China. [6]HKUST Fok Ying Tung Research Institute, Guangzhou, China. [7]The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [8]Laboratorio de Fisicoquímica y Reactividad de Superficies (LaFReS), Instituto de Investigaciones en Materiales, Universidad Nacional Autónoma de México, Coyoacán, CDMX 04510, México. [9]Institute of Organic Chemistry, Justus-Liebig University, Heinrich-Buff-Ring 17, D-35392 Giessen, Germany. [10]Department of Organic Chemistry, Igor Sikorsky Kiev Polytechnic Institute, pr. Pobedy 37, 03056 Kiev, Ukraine. [11]Institute for Molecular Engineering, University of Chicago, Chicago, Illinois 60637, USA. [12]Department of Physics, Stanford University, Stanford, California 94305, USA.
*These authors contributed equally to this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

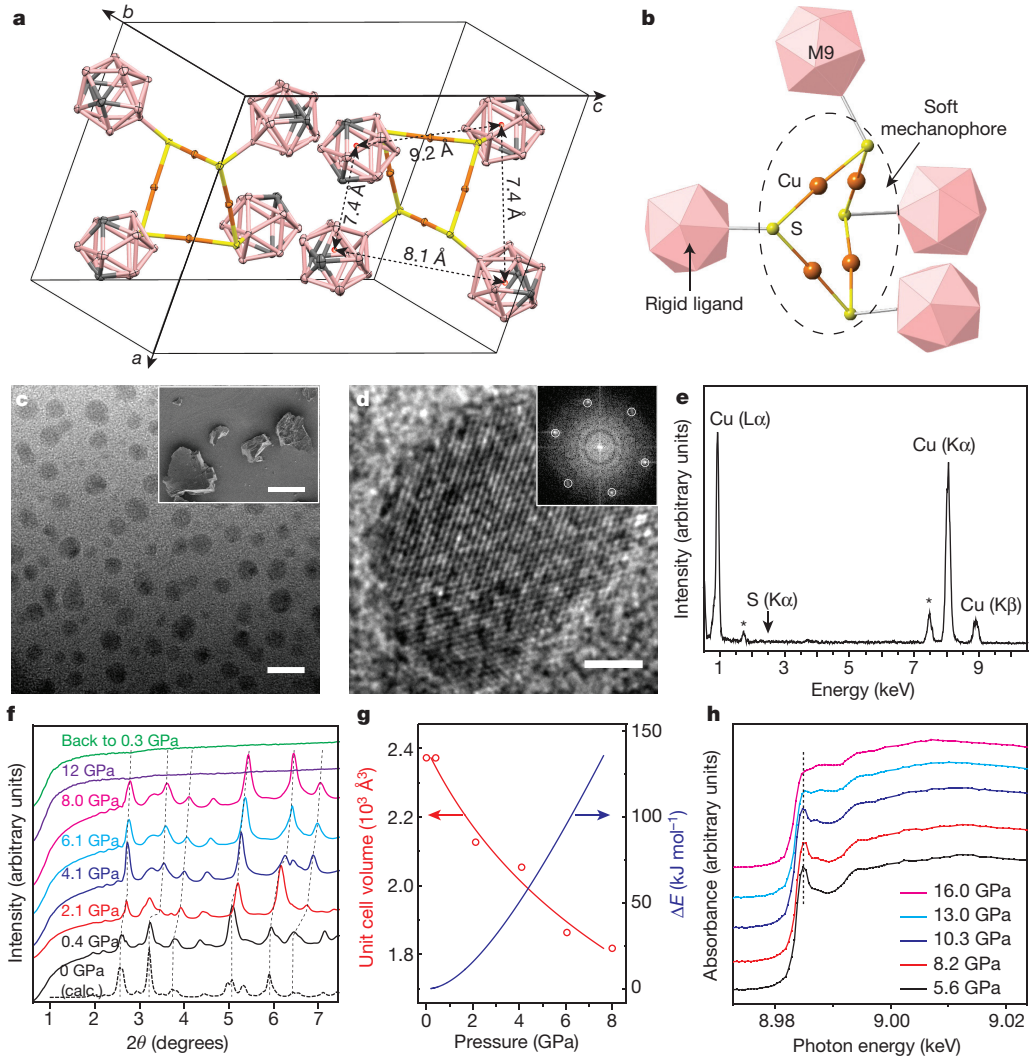


Figure 1 | Redox reaction in Cu-S-M9 under hydrostatic pressure. **a**, Unit cell of Cu-S-M9. Atoms are represented by their thermal ellipsoids at the 50% probability level. Copper, sulfur, carbon and boron atoms are represented by red, yellow, grey and pink ellipsoids, respectively. Hydrogen atoms and interstitial solvent (toluene) molecules are omitted for clarity. **b**, Cu-S-M9 molecule showing the Cu_4S_4 mechanophore surrounded by M9 ligands, represented by polyhedra. **c**, TEM image of nanoclusters formed after compression to 12 GPa. Scale bar, 100 nm. Inset, scanning electron microscopy image of Cu-S-M9 crystals before compression. Scale bar, 50 μm. **d**, High-resolution TEM image of a single copper nanocluster. Scale bar, 2 nm. Inset, Fourier transform pattern. **e**, EDS of the clusters shown in **c** and **d**. The arrow points to the position of the S K-edge. The asterisks mark background signals of Ni and Si. **f**, *In situ* XRD patterns of Cu-S-M9 from 0.4 GPa to 12 GPa and back to 0.3 GPa. $\lambda = 0.4959$ Å. The bottom dashed line shows the ambient XRD pattern calculated from its crystal structure. The vertical dashed lines mark the positions of six selected peaks used to determine the lattice parameters (see Methods and Supplementary Table 5). **g**, Unit cell volume (red, left axis) and total energy change (blue, right axis) as functions of pressure. Open circles and solid lines represent the experimentally derived data and the fitting calculation using a third-order Birch–Murnaghan equation of state, respectively. **h**, *In situ* XAS at different pressures. The dashed line marks the position of the Cu(I) peak.

the characteristic Cu(I) peak (8,985 eV, ref. 23) decreases at 8–13 GPa (Fig. 1h), consistent with the pressure range of structural transformation. Although the decreased Cu(I) peak intensity can be attributed to formation of Cu(0) or Cu(II), *ex situ* X-ray photoelectron spectroscopy data (Extended Data Fig. 4) exclude the formation of Cu(II). These results show that pressure higher than 8 GPa drives the reduction of Cu(I) to form metallic Cu(0).

DFT computations show that the hydrostatic pressure leads to relative motion of the M9 ligands, which in turn change the bonding in the Cu_4S_4 core. We computed the equilibrium structures of Cu-S-M9 at 2–12 GPa (Fig. 2a–d) using the generalized gradient approximation (GGA) level of DFT. The GGA provided a good representation of the system under pressure even without inclusion of dispersion interactions (Methods and Extended Data Fig. 7).

First, the DFT results confirm the large compressibility difference between the Cu_4S_4 mechanophore and M9 ligands. The bond lengths in M9 decrease slightly (<0.1 Å) between ambient pressure and 12 GPa (black and yellow, Fig. 2e), reflecting the rigidity of the ligand in this pressure range. The compression decreases the spacing between the M9 ligands. For example, the distance between the adjacent $M9_1$ and $M9_2$ ligands decreases by >1 Å (red, Fig. 2e). Correspondingly, the distance between the two sulfur atoms connected with $M9_1$ and $M9_2$, S1 and S2, decreases by 0.8 Å (blue, Fig. 2e). These changes are at least one order of magnitude larger than the bond length changes within M9 groups, demonstrating the large compressibility difference between the mechanophore and the ligands. The S2–S3 and $M9_2$–$M9_3$ distances show similar changes (Extended Data Fig. 8).

Second, the computed structures show that the Cu_4S_4 core is anisotropically deformed by the hydrostatic pressure. The system accommodates the reduced S1–S2 distance by bending the S1–Cu1–S2 bond angle from 176° at ambient pressure to 104° at 12 GPa (green, Fig. 2e). The S2–Cu2–S3 bond angle shows the same trend, while

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

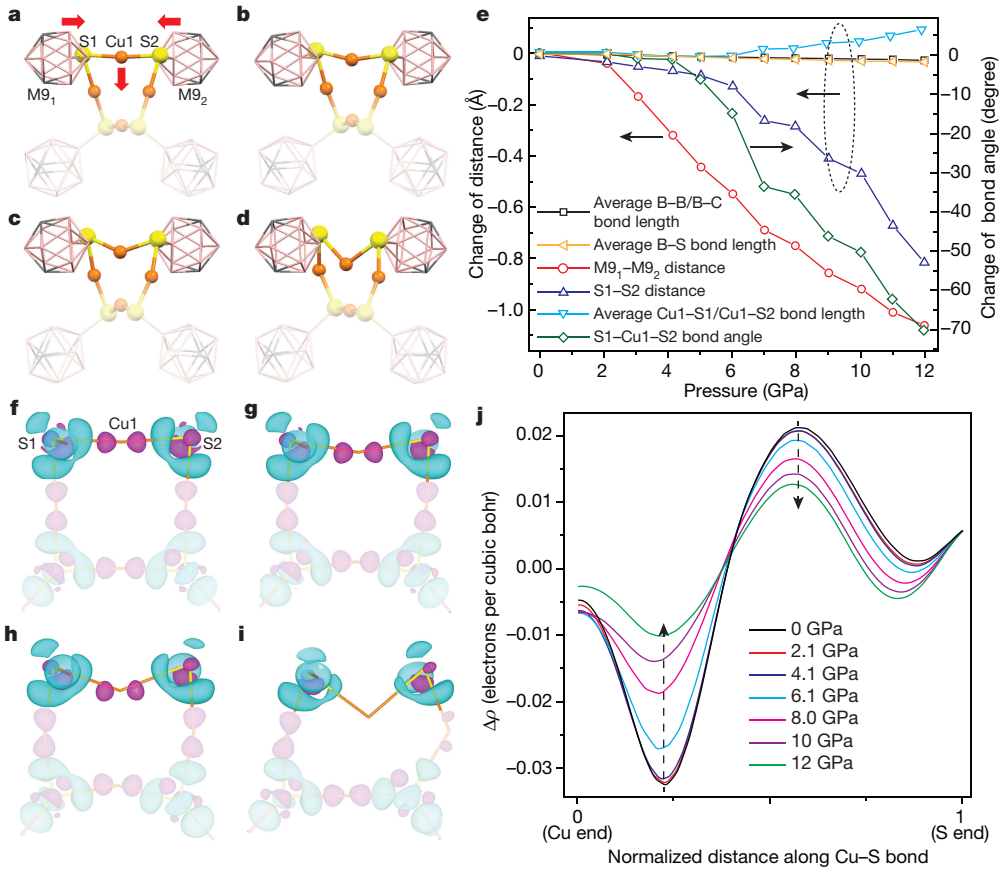


Figure 2 | Modelling the atomic and electronic structures of Cu-S-M9. **a**–**d**, DFT-computed structures of Cu-S-M9 at 2.1 GPa (**a**), 6.1 GPa (**b**), 8.0 GPa (**c**) and 12 GPa (**d**). The red arrows denote the directions of relative motion of copper, sulfur and M9. **e**, Changes of distances (left axis) and bond angle (right axis) as functions of pressure. **f**–**i**, Isosurfaces of $\Delta\rho$ at 2.1 GPa (**f**), 6.1 GPa (**g**), 8.0 GPa (**h**) and 12 GPa (**i**). Cyan and magenta represent isovalues of ±0.012 electrons per cubic bohr, respectively. The carboranes are omitted for clarity. **j**, Averaged line profiles of $\Delta\rho$ along Cu1–S1 and Cu1–S2 bonds at different pressures. The dashed arrows mark the positions of $\Delta\rho$ maxima near sulfur and minima near copper, with the arrow pointing in the direction of increasing pressure.

changes in the two other S–Cu–S bond angles are smaller (Extended Data Fig. 8).

Interestingly, the bending of the S1–Cu1–S2 angle substantially stretches the Cu1–S1 and Cu1–S2 bonds by up to 4% between 7 GPa and 12 GPa (cyan, Fig. 2e), suggesting weakening of the bond strength in this pressure range. The onset pressure is within 1 GPa of the reaction onset pressure observed by *in situ* XRD and XAS. This is due to emergence of nodal planes in the highest occupied molecular orbital (HOMO) across the Cu–S bonds (Extended Data Fig. 9), transforming the HOMO into an antibonding orbital. We note that although equilibrium structures can be computed up to 12 GPa, experimentally the system is sufficiently destabilized above 8 GPa, and reaction takes place spontaneously at room temperature.

The anisotropic deformation of the mechanophore leads to electron transfer from sulfur to copper. Figure 2f–i depicts the electron density difference ($\Delta\rho$), defined as the electron density of the molecule minus the superposition of electron densities of isolated atoms, from ambient pressure to 12 GPa. At ambient pressure, the negative (magenta) and positive (cyan) $\Delta\rho$ near Cu1 and S1/S2 is consistent with the valence of the elements (+1 for Cu and −2 for S). As the pressure increases, both the negative and positive $\Delta\rho$ isosurfaces near Cu1 and S1/S2 shrink in size. At 12 GPa, the electron density near Cu1 is within 0.01 electron per cubic bohr of the electron density of elemental Cu(0). The minima and maxima of $\Delta\rho$ along the Cu–S bonds decrease by 70% and 40% respectively from ambient pressure to 12 GPa (Fig. 2j). These computational results indicate that as the S–Cu–S bond angle is bent, electron density is transferred from S to Cu, resulting in the reduction of Cu(I), consistent with the experimental observation of elemental copper nanoclusters.

The redox reaction in Cu-S-M9 demonstrates that the relative motion of M9 ligands enables anisotropic deformation and reactivity of the Cu–S mechanophore. In contrast, in metal–organic chalcogenides (MOCs) where the rigid ligands contact each other, the relative motion of the ligands is sterically hindered, prohibiting the deformation and reactivity of the metal–chalcogen mechanophore. We demonstrate this 'steric blockage' scenario (Extended Data Fig. 1) in copper(I) adamantane-1-thiolate (Cu-S-Ada, Fig. 3a). The mechanophore, a Cu–S nanowire with a three-atom cross-section, is surrounded by a ligand shell consisting of adamantyl groups. We note that unlike Cu-S-M9, here the adamantyl groups are in van der Waals contact[16,24].

In this case, the sterically hindered ligand shell impedes relative movement of the adamantyl groups and deformation of the Cu–S mechanophore. Equilibrium structures computed by DFT (Fig. 3b) shows that the adamantyl groups are rigid from ambient pressure to 20 GPa, with <0.01 Å change of the average C–C bond lengths (black, Fig. 3c). During the compression process, the decrease of adjacent adamantyl distances (red, Fig. 3c) is 40% less than the maximal M9–M9 distance change in Cu-S-M9, indicating much smaller relative motion. Correspondingly, the average S–S distance and S–Cu–S bond angle changes by about 3% and 13% respectively from ambient pressure to 20 GPa (blue and green, Fig. 3c), substantially smaller than the changes in Cu-S-M9 (about 20% and 40%, respectively, from ambient pressure to 12 GPa). The average Cu–S bond length slightly decreases (cyan, Fig. 3c) at 20 GPa, in clear contrast to the stretching of Cu–S bonds in Cu-S-M9 at 12 GPa and suggesting no bond weakening.

DFT computations indicate a much smaller electron density shift in Cu-S-Ada than in Cu-S-M9. The positive and negative electron

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

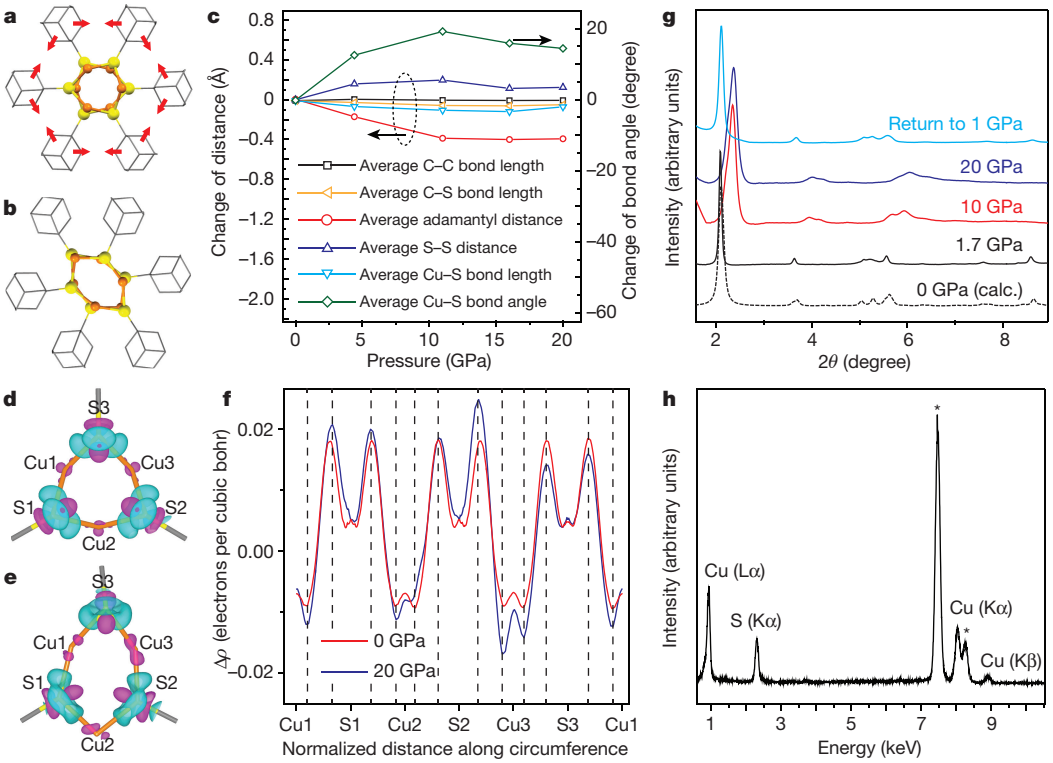


Figure 3 | Sterically impeded reactivity in Cu-S-Ada. a, Ambient-pressure structure of Cu-S-Ada, viewed along the nanowire elongation direction. Red arrows denote the direction of motion for the adamantyl groups under pressure. Copper and sulfur atoms are represented by red and yellow spheres, and the adamantyl groups are represented by the capped stick model. **b**, DFT-computed structure at 20 GPa. **c**, Changes of distances (left axis) and bond angle (right axis) as functions of pressure. **d**, **e**, $\Delta\rho$ isosurfaces at ambient pressure (**d**) and at 20 GPa (**e**). Cyan and magenta represent isovalues of $\pm$0.009 electrons per cubic bohr, respectively. Adamantyl groups are omitted for clarity. **f**, $\Delta\rho$ line profiles along Cu–S bonds on the circumference of the Cu–S nanowire core. The dashed lines mark the positions of $\Delta\rho$ minima near copper atoms and maxima near sulfur atoms. **g**, *In situ* XRD from 1.7 GPa to 20 GPa and back to 1 GPa. $\lambda = 0.4959$ Å. The bottom dashed line shows the XRD pattern calculated from its ambient crystal structure. **h**, EDS of Cu-S-Ada after compression to 20 GPa. The asterisk marks the Ni peak from the TEM grid.

density differences near sulfur and copper atoms persist from ambient pressure up to 20 GPa (Fig. 3d and e). Figure 3f depicts the line profile of $\Delta\rho$ along the Cu–S bonds. It can be seen that the $\Delta\rho$ maxima near S3 decrease by 0.004 electrons per cubic bohr from ambient pressure to 20 GPa, much smaller than the change (about 40%) in Cu-S-M9. On the other hand, the absolute values of $\Delta\rho$ maxima/minima near the other S and Cu atoms increase by up to 0.008 electrons per cubic bohr, indicating an electron density shift from S to Cu in contrast to the S-to-Cu electron transfer observed in Cu-S-M9. These computational results suggest that the reduction of Cu(I) by sulfur, observed in Cu-S-M9, would be hindered in Cu-S-Ada, and we predict that Cu-S-Ada remains mechanochemically inert up to 20 GPa.

This prediction was verified by experiments. *In situ* XRD (Fig. 3g) shows that the diffraction peaks attributed to Cu-S-Ada shift smoothly from ambient pressure to 20 GPa, indicating no amorphization in this pressure range. Upon releasing the pressure, the ambient XRD pattern is fully recovered, showing that the deformation is elastic. EDS of a sample compressed to 20 GPa (Fig. 3h) reveals both copper and sulfur, consistent with the composition of the pristine Cu-S-Ada[16]. The nanowire morphology of the sample is also preserved after compression (Extended Data Fig. 10). No elemental copper clusters could be observed by TEM. These computational and experimental results corroborate that the sterically hindered adamantyl ligand shell in Cu-S-Ada impedes the anisotropic deformation and redox reaction in the Cu–S core. Importantly, the steric control of reactivity does not exist in the thermochemistry of these compounds, as both Cu-S-M9 and Cu-S-Ada yield cuprous sulfide under the same pyrolysis conditions (Extended Data Fig. 5).

The mechanochemistry based on relative motion of rigid ligands can be extended to other molecular architectures. We examined the mechano-reactivity of two different one-dimensional silver diamondoidthiolate structures under hydrostatic pressure. The model systems consisted of two chemically similar, yet sterically different compounds, silver(I) diamantane-1-thiolate (Ag-S-*m*Dia, where *m* denotes the medial position of the thiol group) and silver(I) adamantane-1-thiolate (Ag-S-Ada)[16]. Both structures have two single-atom Ag-S chains packed together via van der Waals interaction and surrounded by diamondoid sidegroups (Fig. 4a and b, Supplementary Table 1, Extended Data Fig. 2). In Ag-S-*m*Dia the two diamantyl groups reside on the same side of the Ag–S chain, adopting a *cis* configuration (Fig. 4a). In contrast, in Ag-S-Ada the two adamantyl groups reside on the opposite side of the Ag-S chain, taking up a *trans* configuration (Fig. 4b). We postulate that in the *cis* configuration, the relative sliding of the two Ag–S chains may occur perpendicular to the chain elongation direction (red arrows, Fig. 4a). In the *trans* configuration steric repulsion between adamantyl groups in adjacent chains would impede relative motion.

DFT computations show that the Ag–S mechanophore in Ag-S-*m*Dia sustains much larger deformation than Ag-S-Ada under hydrostatic pressure. The chain alignment angle, defined by sulfur atoms in the two adjacent chains as shown in Fig. 4a, increases by >10% in Ag-S-*m*Dia from ambient pressure to 20 GPa (blue, Fig. 4c). In contrast, the chain alignment angle in Ag-S-Ada decreases by approximately 2% in the same pressure range (red, Fig. 4c). Instead of shearing the mechanophore, the Ag-S-Ada system absorbs the stress by buckling the Ag–S chains.

This displacement in Ag-S-*m*Dia chains results in an electron density shift towards silver. At ambient pressure, negative and positive $\Delta\rho$ isosurfaces exist near all Ag and S atoms of Ag-S-*m*Dia (upper panel of Fig. 4d), consistent with their respective valences (+1 for Ag and −2 for S). At 20 GPa, however, the $\Delta\rho$ near half of the Ag atoms (Ag1 and Ag2, lower panel of Fig. 4d) decreases below 0.01 electrons

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

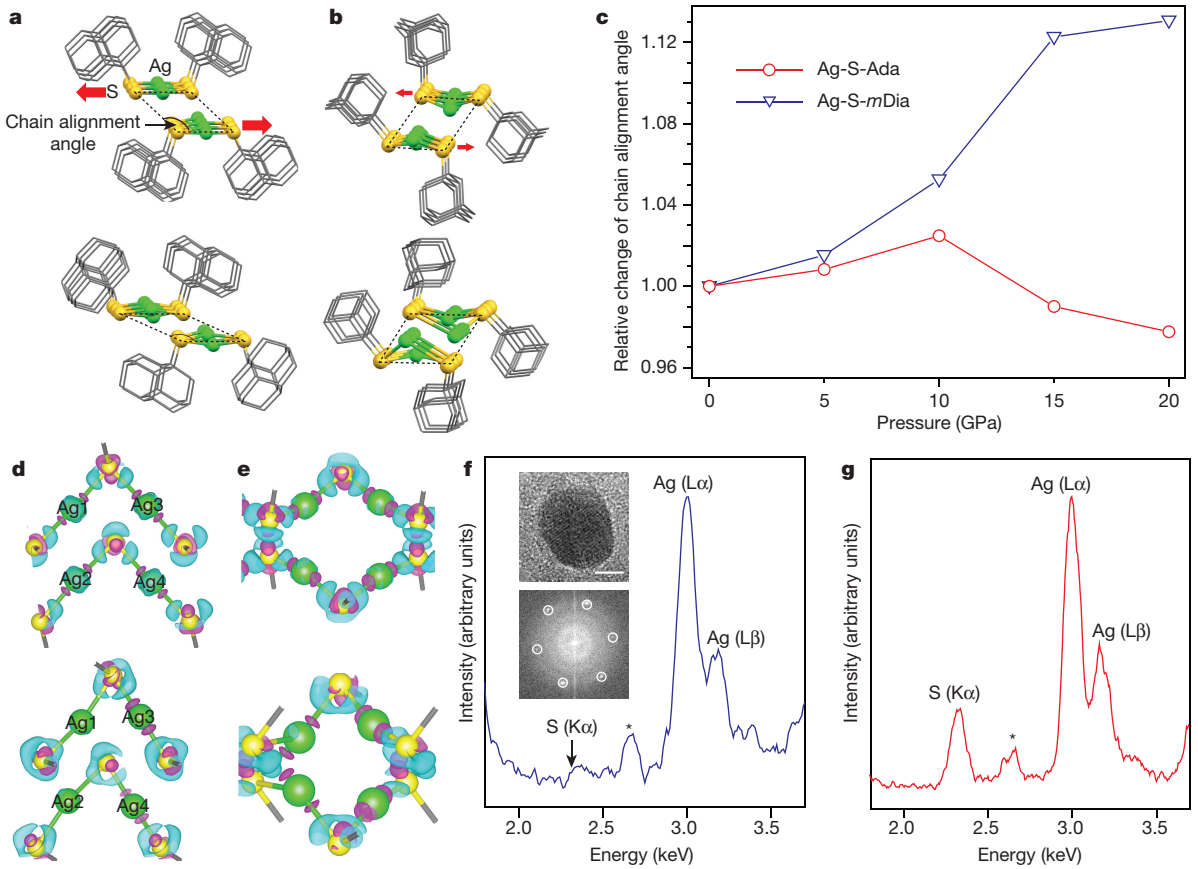


Figure 4 | Sterically controlled deformation and reactivity in one-dimensional structures. a, **b**, Atomic structures of Ag-S-*m*Dia (**a**) and Ag-S-Ada (**b**) at ambient pressure (top) and 20 GPa (bottom). Ambient-pressure and high-pressure structures are determined by single-crystal XRD and DFT computation, respectively. The structures are viewed along the chain elongation directions. Silver and sulfur are represented by green and yellow spheres, and the diamondoids are represented by capped stick models. The red arrows denote the directions of motion for the two adjacent Ag–S chains. **c**, Relative changes of chain alignment angles as functions of pressure for Ag-S-Ada and Ag-S-*m*Dia. **d**, **e**, $\Delta\rho$ isosurfaces for Ag-S-*m*Dia (**d**) and Ag-S-Ada (**e**) at ambient (top) and 20 GPa (bottom) pressures. Cyan and magenta represent isovalues of $\pm$0.01 electrons per cubic bohr, respectively. **f**, EDS of Ag-S-*m*Dia after 20 GPa compression. Top inset, TEM image of a silver nanocluster. Scale bar, 5 nm. Bottom inset, Fourier transform pattern of the TEM image. **g**, EDS of Ag-S-Ada after 20 GPa compression. The asterisks in **f** and **g** mark background Cl signals from solvent residue.

per cubic bohr. The magnitude of this decrease in $\Delta\rho$ near Ag1 and Ag2 is comparable to that near Cu1 of Cu-S-M9 at 12 GPa, and suggests the reduction of Ag(I) to Ag(0) at 20 GPa. On the other hand, although the Ag–S chain in Ag-S-Ada is buckled by hydrostatic compression, the negative $\Delta\rho$ adjacent to all Ag atoms persists from ambient pressure to 20 GPa (Fig. 4e), suggesting that the redox reaction is blocked within this pressure range.

The different mechano-reactivity predicted by DFT is confirmed by the experiments. TEM of a Ag-S-*m*Dia sample after compression to 20 GPa (upper inset of Fig. 4f) reveals crystalline nanoclusters with lattice spacing of 2.4 Å, consistent with the [111] spacing of silver. EDS (Fig. 4f) shows that these clusters consist exclusively of silver and are free of sulfur, thus confirming the formation of elemental silver by hydrostatic compression. In contrast, no silver clusters can be observed in Ag-S-Ada samples after compression to 20 GPa. Although large particles are observed (Extended Data Fig. 10), these particles have no identifiable lattice under TEM. Furthermore, EDS reveals that these particles consist of both silver and sulfur (Fig. 4g), consistent with the composition of pristine Ag-S-Ada. We conclude that although the 20 GPa hydrostatic pressure changes the morphology (from nanowire to particles), and possibly the crystallinity, there is no evidence of a redox reaction in this compound. The computational and experimental results demonstrate that steric effects control the deformation and associated reactivity in one-dimensional MOCs, implementing the molecular anvil and steric blockage scenarios with a different reaction mechanism from the bond angle bending in Cu-S-M9.

The molecular anvils offer an atomic-efficient design of mechano-chemical systems with a ligand-to-mechanophore atomic ratio of the order of ten. In contrast, polymeric mechanochemical systems require long chains with about a thousand repeating units supporting the mechanophore[25], with a polymer-to-mechanophore atomic ratio of at least one hundred. These results suggest that rigid and sterically bulky functional groups, which are widely explored in conventional organic synthesis as protecting and directing groups, might also be readily adopted for mechanochemistry, opening new methods towards highly specific mechanosynthesis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 September 2017; accepted 8 January 2018.

1. Gilman, J. J. Mechanochemistry. *Science* **274,** 65–66 (1996).
2. Caruso, M. M. *et al.* Mechanically-induced chemical changes in polymeric materials. *Chem. Rev.* **109,** 5755–5798 (2009).
3. Hickenboth, C. R. *et al.* Biasing reaction pathways with mechanical force. *Nature* **446,** 423–427 (2007).
4. May, P. A. *et al.* Polymer mechanochemistry: techniques to generate molecular force via elongational flows. *Chem. Soc. Rev.* **42,** 7497–7506 (2013).
5. Huang, Z. & Boulatov, R. Chemomechanics: chemical kinetics for multiscale phenomena. *Chem. Soc. Rev.* **40,** 2359–2384 (2011).
6. Diesendruck, C. E. *et al.* Mechanically triggered heterolytic unzipping of a low-ceiling-temperature polymer. *Nat. Chem.* **6,** 623–628 (2014).
7. Piermattei, A., Karthikeyan, S. & Sijbesma, R. P. Activating catalysts with mechanical force. *Nat. Chem.* **1,** 133–137 (2009).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

8. Klukovich, H. M. *et al.* A backbone lever-arm effect enhances polymer mechanochemistry. *Nat. Chem.* **5,** 110–114 (2013).
9. Chen, Y. *et al.* Mechanically induced chemiluminescence from polymers incorporating a 1,2-dioxetane unit in the main chain. *Nat. Chem.* **4,** 559–562 (2012).
10. Davis, D. A. *et al.* Force-induced activation of covalent bonds in mechanoresponsive polymeric materials. *Nature* **459,** 68–72 (2009).
11. Wang, J., Kouznetsova, T. B. & Craig, S. L. Single-molecule observation of a mechanically activated *cis*-to-*trans* cyclopropane isomerization. *J. Am. Chem. Soc.* **138,** 10410–10412 (2016).
12. Chen, Z. *et al.* Mechanochemical unzipping of insulating polyladderene to semiconducting polyacetylene. *Science* **357,** 475–479 (2017).
13. Li, W. & Gräter, F. Atomistic evidence of how force dynamically regulates thiol/disulfide exchange. *J. Am. Chem. Soc.* **132,** 16790–16795 (2010).
14. Dopieralski, P. *et al.* The Janus-faced role of external forces in mechanochemical disulfide bond cleavage. *Nat. Chem.* **5,** 685–691 (2013).
15. Akbulatov, S. *et al.* Experimentally realized mechanochemistry distinct from force-accelerated scission of loaded bonds. *Science* **357,** 299–303 (2017).
16. Yan, H. *et al.* Hybrid metal–organic chalcogenide nanowires with electrically conductive inorganic core through diamondoid-directed assembly. *Nat. Mater.* **16,** 349 (2017).
17. Gamba, Z. & Powell, B. M. The condensed phases of carboranes. *J. Chem. Phys.* **105,** 2436 (1996).
18. Schwertfeger, H., Fokin, A. A. & Schreiner, P. R. Diamonds are a chemist's best friend: diamondoid chemistry beyond adamantane. *Angew. Chem. Int. Ed.* **47,** 1022–1036 (2008).
19. Gunawan, M. A. *et al.* Diamondoids: functionalization and subsequent applications of perfectly defined molecular cage hydrocarbons. *New J. Chem.* **38,** 28–41 (2014).
20. Castner, D. G., Hinds, K. & Grainger, D. W. X-ray photoelectron spectroscopy sulfur 2p study of organic thiol and disulfide binding interactions with gold surfaces. *Langmuir* **12,** 5083–5086 (1996).
21. Schneider, S., Dzudza, A., Raudaschl-Sieber, G. & Marks, T. J. Copper(I) *tert*-butylthiolato clusters as single-source precursors for high-quality chalcocite thin films: precursor chemistry in solution and the solid state. *Chem. Mater.* **19,** 2768–2779 (2007).
22. Birch, F. Finite elastic strain of cubic crystals. *Phys. Rev.* **71,** 809–824 (1947).
23. Jang, Y. J. *et al.* Tree branch-shaped cupric oxide for highly effective photoelectrochemical water reduction. *Nanoscale* **7,** 7624–7631 (2015).
24. Nowacki, W. Die Krystallstruktur von Adamantan (symm. *Tri-cyclo-decan). Helv. Chim. Acta* **28,** 1233–1242 (1945).
25. May, P. A. *et al.* Is molecular weight or degree of polymerization a better descriptor of ultrasound-induced mechanochemical transduction? *ACS Macro Lett.* **5,** 177–180 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Beavers, J. Yan and S. Teat from the Advanced Light Source for help with XRD measurements, and C. Park and D. Popov from the Advanced Photon Source for XAS measurements. This work was supported by the Department of Energy, Office of Basic Energy Sciences, Division of Materials Science and Engineering, under contracts DE-AC02-76SF00515 and DE-FG02-06ER46262. D.P. acknowledges support from Hong Kong Research Grants Council (project number ECS-26305017), the National Natural Science Foundation of China (project number 11774072) and the Alfred P. Sloan Foundation through the Deep Carbon Observatory. D.S.-I. acknowledges support from PAPIIT IA203116/27 and CONACYT FC-2015-2/829. This research used resources of the Advanced Light Source, which is a US Department of Energy (DOE) Office of Science User Facility under contract DE-AC02-05CH11231. This research also used resources of the Advanced Photon Source, a DOE Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract DE-AC02-06CH11357. Portions of this work were performed at the Stanford Nano Shared Facilities, supported by the National Science Foundation under award ECCS-1542152. The computational work used resources at the Stanford Research Computing Center, the Research Computing Center at the University of Chicago, and the Deep Carbon Observatory computer cluster.

Author Contributions H.Y., W.L.M., Z.-X.S. and N.A.M. conceived the idea. H.Y., J.N.H. and D.S.-I. synthesized the crystals and solved their structures. H.Y., F.Y. and Y.L. carried out the high-pressure experiments. H.Y., D.P. and G.G. performed the DFT computations. H.Y. and F.H.L. performed the *ex situ* characterizations. J.E.P.D., R.M.K.C., B.A.T., A.A.F. and P.R.S. provided the diamondoids and synthesized their derivatives. H.Y. and N.A.M. wrote the paper. All authors contributed to the discussion and revision of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to N.A.M. (nmelosh@stanford.edu).

Reviewer Information *Nature* thanks D. Braga, S. James and L. Yan for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Synthesis and structural determination of MOCs. All commercially available chemicals were purchased from Sigma Aldrich and used without further purification. Anhydrous chemicals were used when possible. Synthesis of Cu-S-Ada, Ag-S-Ada and Ag-S-*m*Dia was described previously[16]. Cu-S-M9 was synthesized in an argon-filled glove box. 20 mM solutions of *m*-carborane-9-thiol in toluene and copper sulfate in ethylene glycol were layered in a glass vial, with 1:1 volumetric ratio and the toluene phase on top. The mixture was heated at 70 °C for a week. Upon cooling to room temperature, colourless crystals appear at the toluene/ethylene glycol interface. The crystals were harvested by vacuum filtration, washed with ethanol and dried under vacuum.

Single-crystal XRD of Cu-S-M9 and Ag-S-Ada was performed using a Bruker D8 Venture diffractometer, equipped with a MoK_α source (wavelength $\lambda = 0.7107$ Å) and a Photon 100 complementary metal-oxide-semiconductor detector. Single-crystal XRD of Ag-S-*m*Dia was measured with synchrotron radiation ($\lambda = 0.7749$ Å) at Beamline 11.3.1, Advanced Light Source, Lawrence Berkeley National Laboratory, with a Bruker D8 diffractometer and a Bruker AXS APEXII CCD detector. Crystals were coated with Paratone-N oil, attached to Kapton loops, and mounted on the diffractometer. Frames were collected using ω and/or ψ scans and the unit-cell parameters were refined against all data. Data were integrated and corrected for Lorentz and polarization factors, as well as the absorption effects using SAINT 8.27b and SADABS software integrated in the APEX2 suite[26]. Space-group assignments were based upon systematic absences, *E*-statistics, agreement factors for equivalent reflections, and successful refinement of the structure. The structures were solved by direct methods and expanded through successive difference Fourier maps using SHELXS-97. They were refined against all data using SHELXTL and OLEX2 software[27–29]. Hydrogen atoms were inserted at idealized positions and refined using a riding model with an isotropic thermal parameter 1.2 times that of the attached atom. Thermal parameters for all non-hydrogen atoms were refined anisotropically. Details of the data quality and a summary of the residual values of the refinements are listed in Supplementary Table 1.

High pressure experiments. Compression and *in situ* XRD/XAS were performed using a diamond anvil cell with a 500-μm diamond culet. The sample chamber was created by drilling a 150-μm hole in a pre-indented stainless steel gasket. Two types of experiments were performed. In the first, the chamber was filled with approximately a few thousand crystals in the size range 1–10 μm, and no pressure medium was used. In the second scenario, a single crystal with size approximately 100 μm was loaded in the chamber together with neon as the pressure medium. In both cases we observed formation of elemental copper after compression beyond 12 GPa. The pressure was monitored with the photoluminescence of a 5-μm ruby ball loaded together with the samples. Pressure was increased and decreased gradually in steps of a few hundred megapascals, and roughly at a rate of 1 GPa min^{-1}. Pressure values reported throughout the paper have 5% uncertainty[30].

In situ powder XRD was performed at Beamline 12.2.2, Advanced Light Source, Lawrence Berkeley National Laboratory. A monochromatic beam ($\lambda = 0.4959$ Å) was used. Samples were measured with the Debye–Scherrer geometry. The Debye rings were integrated using the FIT2D software[31]. To derive the lattice parameters from the diffraction patterns, the lattice spacings corresponding to six diffraction peaks with known Miller indices (vertical dashed lines in Fig. 1f, see also Supplementary Table 5) were calculated at different pressures. Note that in triclinic crystals

$$\frac{1}{d_{hkl}^2} = \frac{1}{V^2}(S_{11}h^2 + S_{22}k^2 + S_{33}l^2 + 2S_{12}hk + 2S_{23}kl + 2S_{13}hl)$$

where d is the lattice spacing, hkl is the Miller index, a, b, c, α, β and γ are the lattice parameters, V is the unit cell volume, and

$$S_{11} = b^2c^2\sin^2\alpha$$

$$S_{22} = a^2c^2\sin^2\beta$$

$$S_{33} = a^2b^2\sin^2\gamma$$

$$S_{12} = abc^2(\cos\alpha\cos\beta - \cos\gamma)$$

$$S_{23} = a^2bc(\cos\beta\cos\gamma - \cos\alpha)$$

$$S_{13} = ab^2c(\cos\alpha\cos\gamma - \cos\beta)$$

The lattice parameters can thus be found by numerically solving the nonlinear equation sets.

High pressure XAS experiments were performed at Beamline 16-BMD HPCAT, Advanced Photon Source, Argonne National Laboratory. Monochromatic X-rays focused by Kirkpatrick–Baez mirrors were directed through a symmetric diamond anvil cell. The scattered X-ray was then energy-selected by an analyser and collected by the detector. The entire edge was scanned from 8,970 eV to 9,025 eV with a step size of 0.25 eV. The pre-edge was scanned from 8,980 eV to 8,990 eV with a step size of 0.1 eV.

***Ex situ* characterizations.** TEM imaging and EDS were collected on an FEI Tecnai F20 system operating at 200-kV accelerating voltage. To prepare the TEM samples, the metal gasket containing the post-compression material was unloaded from the diamond anvil cell and sonicated in isopropanol. The supernate was drop-cast on a nickel TEM grid (Ted Pella). The TEM grid was then rinsed in chloroform and dried in vacuum. We note that the sonication or imaging process does not produce metal nanoclusters (Extended Data Fig. 3).

Scanning electron microscopy images were acquired with an FEI Magellan system operating at 2-kV acceleration voltage. Sample powders were gently pressed onto a degenerately doped silicon wafer, and sputter-coated with about 5-nm-thick Au/Pd before imaging.

X-ray photoelectron spectroscopy was measured on a PHI Versaprobe system with monochromated AlK_α source (1,486 eV). Pristine sample powders were pressed onto a degenerately doped silicon wafer. Compressed samples were characterized in the gasket.

Ex situ XRD was performed on a Bruker D8 Venture system equipped with a MoK_α source ($\lambda = 0.7107$ Å) and a complementary metal-oxide-semiconductor detector. Powder samples were filled into borosilicate capillary tubes and measured with the Debye–Scherrer geometry. The Debye rings were integrated using the APEX3 software suite.

DFT computations. DFT computations were performed with the plane-wave basis set, the Perdew–Burke–Ernzerhof (PBE) exchange-correlation functional[32] and the projector augmented wave method, implemented in the Quantum ESPRESSO package[33]. For each pressure, the crystal structure was first optimized using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm. All atomic positions and lattice parameters were optimized. The kinetic energy cutoffs for wavefunctions and electron density were set at 60 and 240 Rydberg (Ry) for the optimization. The convergence threshold for energy, force and pressure were set to 10^{-4} Ry, 10^{-4} Ry per bohr and 10 MPa, respectively. After convergence of the structural optimization, another single-point calculation with higher cutoffs (80 Ry for wavefunctions and 800 Ry for electron density) was performed to determine the electron density, total energy and stress tensor. The off-diagonal elements of the stress tensors are less than 1% of the diagonal elements (Supplementary Tables 3 and 4), confirming the isotropy of the pressure in these computations. The structural and volumetric data were visualized with Mercury[34] and VESTA[35].

Dispersion interactions are important in determining the ambient structures of MOCs[16]. We included dispersion interactions in the DFT computations of the high-pressure structures using nonlocal exchange-correlation functionals including vdW-DF[36], vdW-DF2[37] and vdW-DF-cx[38]. We found that the unit cell volumes of Cu-S-M9 computed with PBE, vdW-DF and vdW-DF2 all agree reasonably well with experimental data, with less than 5% deviation, while vdW-DF-cx consistently underestimates the unit cell volume, probably owing to overestimation of the dispersion interaction. Furthermore, both PBE and vdW-DF robustly reproduce the key structural changes in Cu-S-M9, that is, shrinkage of the S1–S2 distance, bending of the S1–Cu1–S2 bond angle and elongation of the Cu−S bond beyond 7 GPa (Extended Data Fig. 7). These results indicate that the dispersion interaction does not substantially affect the computed high-pressure structures.

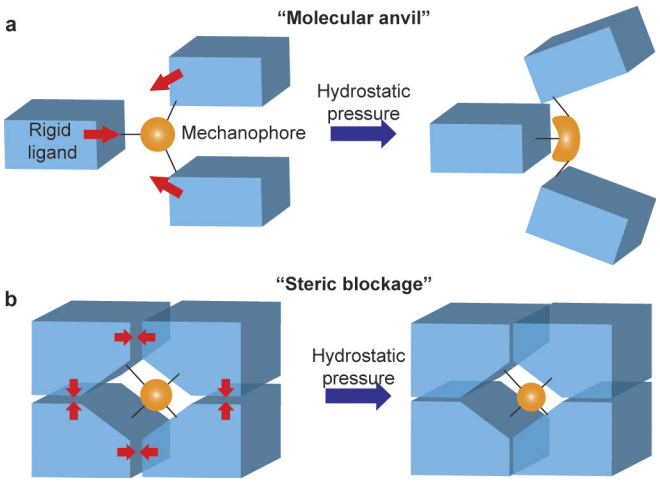
Data availability. Additional data supporting the findings of this study, including raw images, spectra and diffraction patterns, as well as DFT input and output files, are available from the authors. The crystallographic data have been deposited with the Cambridge Crystallographic Data Centre under CCDC-1589700, CCDC-1589701 and CCDC-1589702. These data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif.

26. APEX2, https://www.bruker.com/products/x-ray-diffraction-and-elemental-analysis/single-crystal-x-ray-diffraction/overview/sc-xrd-software/apex3.html (Bruker AXS, 2007).
27. Sheldrick, G. M. *et al.* Crystal structure refinement with *SHELXL*. *Acta Crystallogr. C.* **71,** 3–8 (2015).
28. Sheldrick, G. M. A short history of *SHELX*. *Acta Crystallogr. A* **64,** 112–122 (2008).
29. Dolomanov, O. V. *et al. OLEX2*: a complete structure solution, refinement and analysis program. *J. Appl. Cryst.* **42,** 339–341 (2009).
30. Mao, H.-K. & Mao, W. L. in *Treatise on Geophysics* 231–267 (Elsevier, 2007).
31. Hammersley, A. P. *Internal Report, ESRF98HA01T, FIT2D V9.129 Reference Manual V3.1* (European Synchrotron Radiation Facility (ESRF), 1998).
32. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77,** 3865–3868 (1996).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

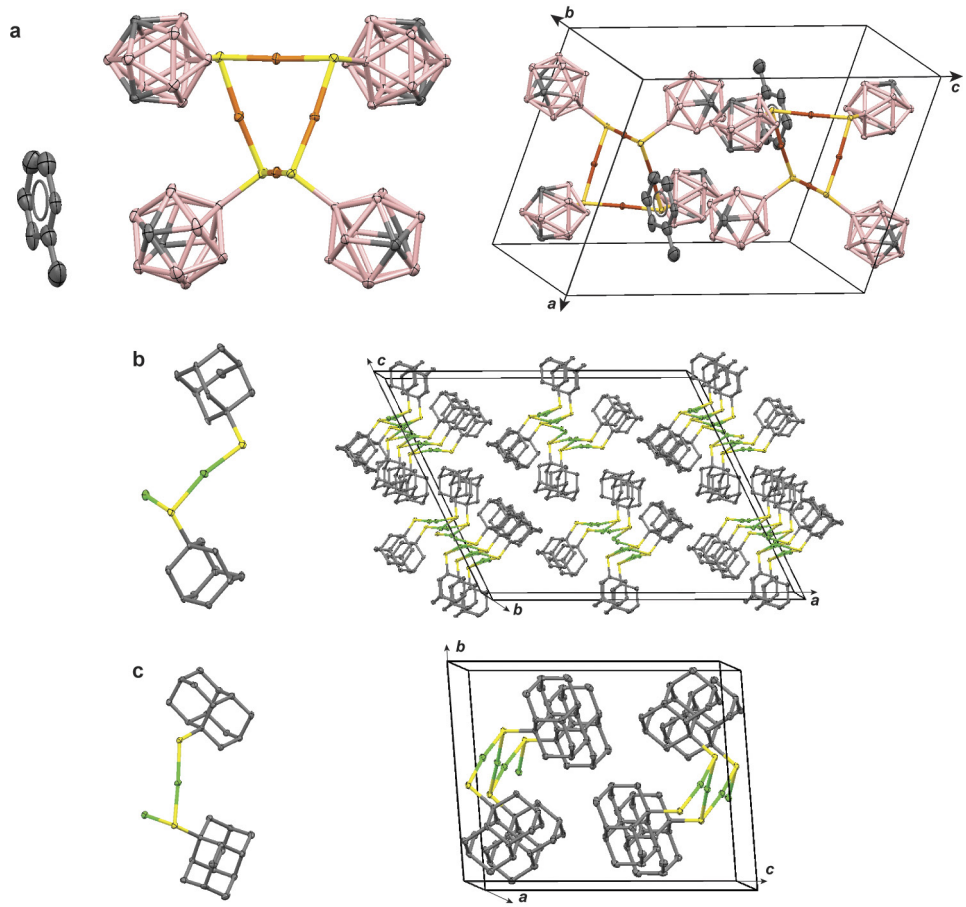
33. Giannozzi, P. *et al.* QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21,** 395502 (2009).
34. Macrae, C. F. *et al.* Mercury: visualization and analysis of crystal structures. *J. Appl. Cryst.* **39,** 453–457 (2006).
35. Momma, K. & Izumi, F. *VESTA 3* for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Cryst.* **44,** 1272–1276 (2011).
36. Dion, M. *et al.* Van der Waals density functional for general geometries. *Phys. Rev. Lett.* **92,** 246401 (2004).
37. Lee, K. *et al.* Higher-accuracy van der Waals density functional. *Phys. Rev. B* **82,** 081101 (2010).
38. Berland, K. *et al.* Assessment of two hybrid van der Waals density functionals for covalent and non-covalent binding of molecules. *J. Chem. Phys.* **146,** 234106 (2017).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



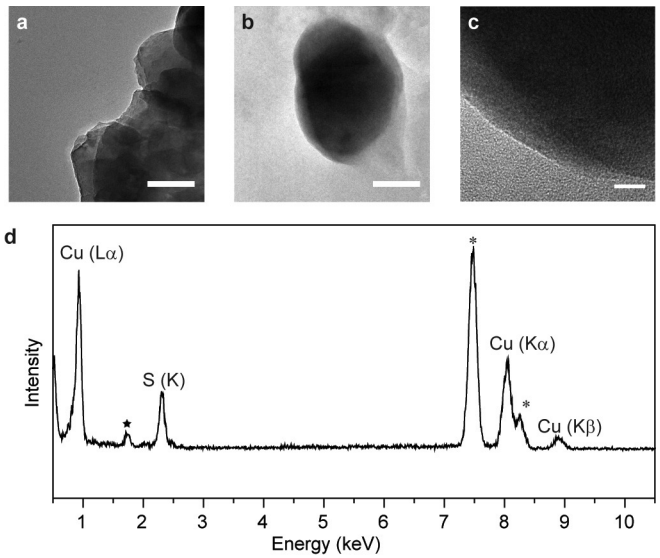
Extended Data Figure 1 | Schematic of the molecular anvil and steric blockage scenarios. a, Relative motions (red arrows) of the rigid ligands under hydrostatic pressure anisotropically deform the mechanophore, leading to reactivity. **b**, Motion of the ligands is blocked by their steric repulsion, protecting the mechanophore from deformation.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



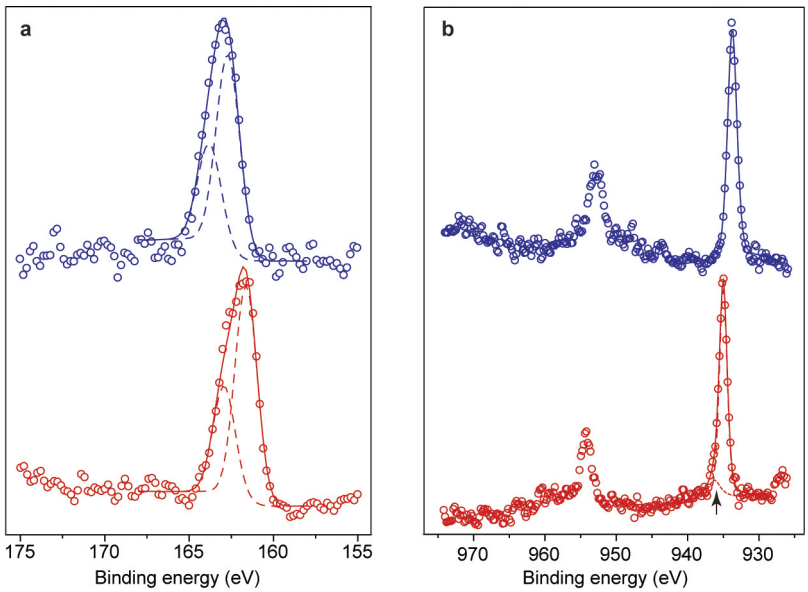
Extended Data Figure 2 | Crystal structures of compounds used in this study. a, Cu-S-M9. **b**, Ag-S-Ada. **c**, Ag-S-*m*Dia. Each panel depicts the asymmetric unit (left) and the unit cell (right). Atoms are represented by their thermal ellipsoids at the 50% probability level. Copper, silver, sulfur, carbon and boron are denoted by red, green, yellow, grey and pink respectively. Hydrogen atoms are omitted for clarity. The toluene molecules (depicted in grey) reside in the interstitial spaces between Cu-S-M9 molecules with a 1:1 molar ratio.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



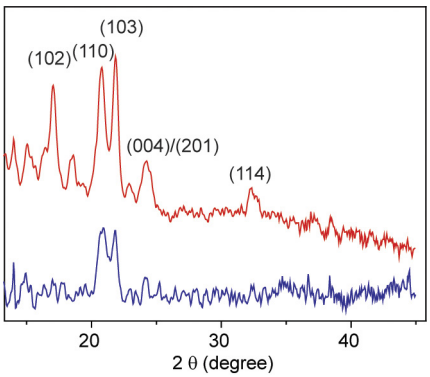
Extended Data Figure 3 | Unreacted Cu-S-M9. a, TEM image of uncompressed Cu-S-M9. **b**, TEM image of Cu-S-M9 after compression to 8 GPa. Scale bars in **a** and **b**, 100 nm. **c**, Zoom-in view of **b**. Scale bar, 10 nm. No inorganic lattice structures are identified in these samples. **d**, Representative EDS of unreacted Cu-S-M9. Both Cu and S peaks can be seen. The asterisks mark Ni peaks originating from the TEM grid. The star marks the position of the Si K-edge, an impurity probably introduced during TEM sample preparation.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



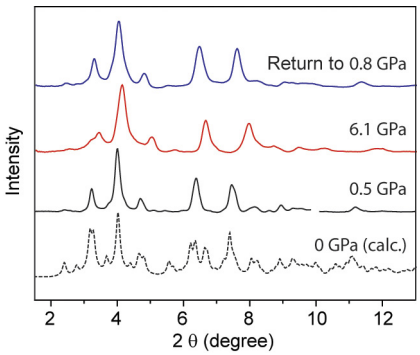
Extended Data Figure 4 | S 2*p* and Cu 2*p* X-ray photoelectron spectroscopy of Cu-S-M9. Red and blue represent uncompressed Cu-S-M9 and sample after compression to 12 GPa, respectively. Discrete dots and solid continuous lines denote experimental data and fitting, respectively. **a**, The S 2*p* spectra are fitted to two Voigt peaks representing $2p_{1/2}$ and $2p_{3/2}$ peaks (dashed lines). The peak position of the uncompressed sample, 162 eV, is characteristic of metal thiolates. The upshift of the binding energy to 163 eV after compression can be attributed to the oxidation of sulfur to form, for example, disulfides. **b**, The Cu 2*p* spectra consist of two peaks, namely $2p_{1/2}$ at about 936 eV and $2p_{3/2}$ at about 954 eV. The absence of satellite peaks excludes the formation of Cu(II) species. Moreover, the $2p_{3/2}$ peak of the uncompressed sample is best fitted by two Voigt peaks (the shoulder peak is denoted by the dashed line and marked by the arrow), characteristic of Cu(I). The $2p_{3/2}$ peak of the compressed sample is best fitted with a single Voigt peak, consistent with Cu(0). These features support the conclusion that copper is in the +1 valence state in the pristine sample, and reduced to the zero valence state upon compression to 12 GPa.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



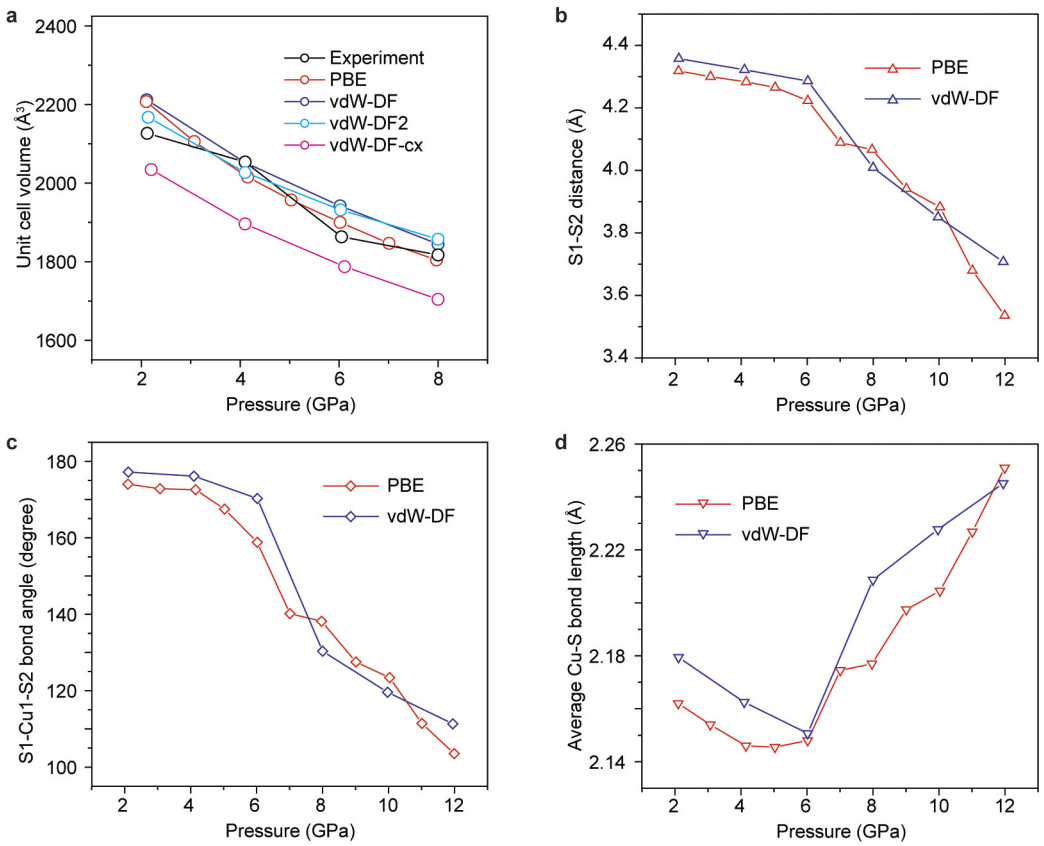
Extended Data Figure 5 | XRD patterns of the pyrolysis products of Cu-S-M9 and Cu-S-Ada. Peaks are registered to β-Cu_2S (chalcocite, PDF Number 00-026-1116). The pyrolysis product of Cu-S-Ada (red) is more crystallized than that of Cu-S-M9 (blue); however, the two strongest peaks, (110) at 20.75° ($d = 1.96$ Å) and (103) at 21.85° ($d = 1.87$ Å) can be clearly seen in both samples. Pyrolysis was carried out at 400 °C in sealed quartz tubes under vacuum. XRD was recorded with a MoK_α source ($\lambda = 0.7107$ Å).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



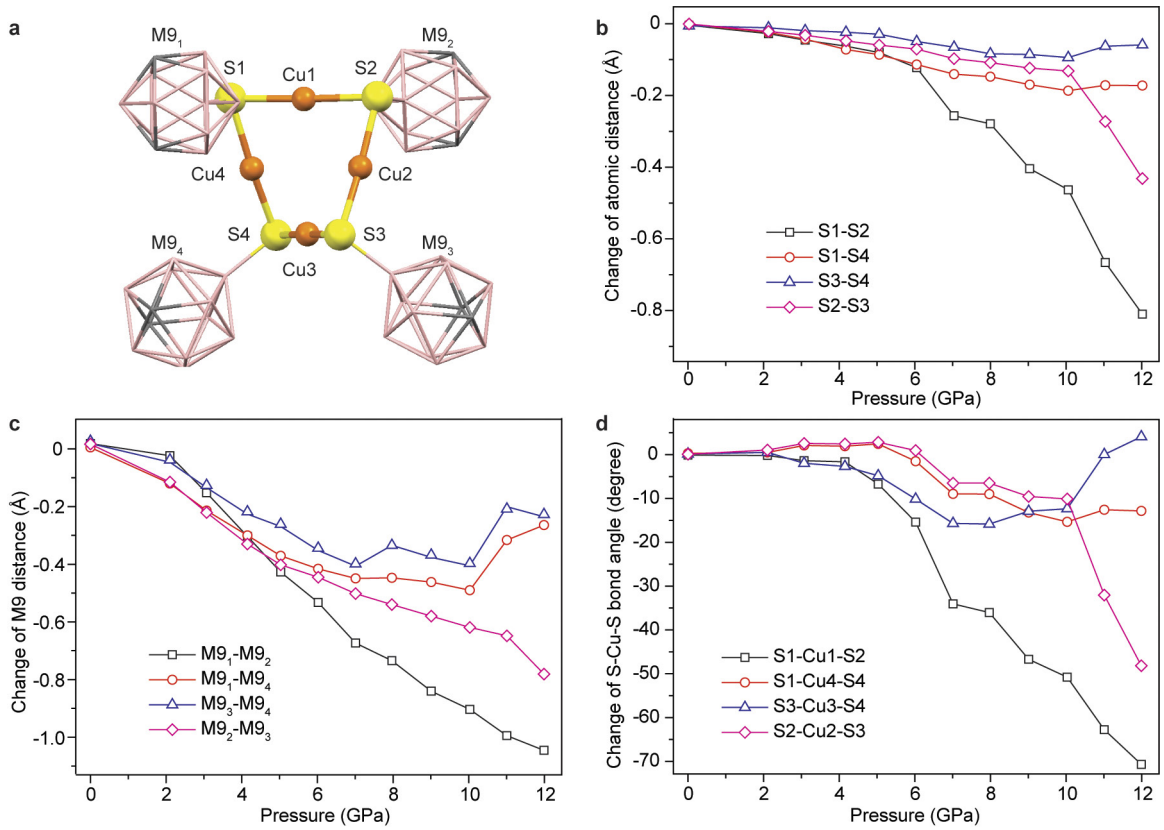
Extended Data Figure 6 | Reversible compression of Cu-S-M9 below 8 GPa. The dashed line shows ambient XRD calculated from the single-crystal structure. The solid lines are *in situ* XRD measured at 0.5 GPa (black), 6.1 GPa (red) and back to 0.8 GPa (blue). $\lambda = 0.6199$ Å.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



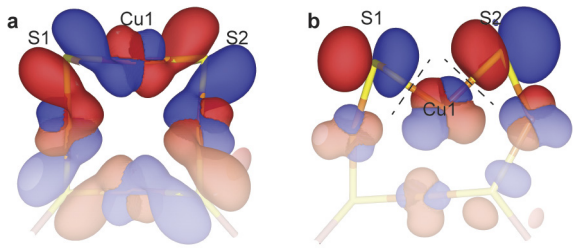
Extended Data Figure 7 | Experimental and computed unit cell volumes and key structural features of Cu-S-M9 at high pressure. a, Unit cell volumes determined by experiment and computed by DFT using different exchange-correlation functionals. **b**–**d**, S1–S2 distance (**b**), S1–Cu1–S2 bond angle (**c**) and average Cu1–S1/Cu1–S2 bond length (**d**) computed using PBE (red) and vdW-DF (blue) functionals. The unit cell volumes computed by PBE, vdW-DF and vdW-DF2 agree well with experimental data, with less than 5% deviation. The vdW-DF-cx functional gave consistently smaller unit cell volume, probably owing to overestimation of the dispersion interaction. Furthermore, PBE and vdW-DF show the same trends in the key structural changes, that is, decreasing S1–S2 distance and S1–Cu1–S2 bond angle, as well as increasing Cu–S bond length beyond 6 GPa. These results show that the van der Waals interactions do not substantially affect the computed high-pressure structures in our systems.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



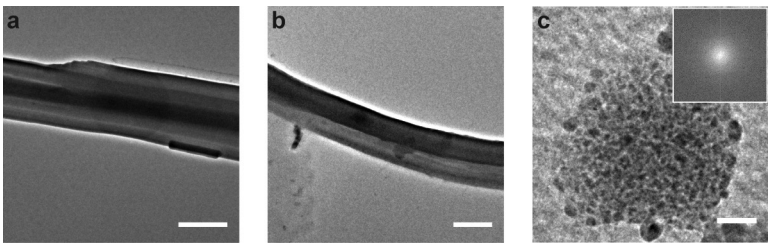
Extended Data Figure 8 | Structural changes of Cu-S-M9 at high pressure. a, Structure of the Cu-S-M9 molecule. **b**–**d**, Changes in S–S distances (**b**), M9–M9 distances (**c**) and S–Cu–S bond angles (**d**) as functions of pressure. All data are extracted from DFT computations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 9 | HOMO of Cu-S-M9, uncompressed and compressed. a, At ambient pressure; **b**, at 12 GPa. The isosurfaces depict the probability density times the sign of the HOMO wavefunction ψ. Red and blue represent isovalues of $\pm 2 \times 10^{-4}$ per cubic bohr, respectively. The dashed lines in the 12 GPa structure mark the nodal planes across Cu–S bonds.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | TEM images of sterically impeded systems. a, **b**, Cu-S-Ada before (**a**) and after (**b**) 20 GPa compression. Scale bars in **a** and **b**, 1 μm. **c**, Ag-S-Ada after 20 GPa compression. Scale bar, 20 nm. Inset, Fourier transform pattern of **c**.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature25749

Palladium-catalysed electrophilic aromatic C–H fluorination

Kumiko Yamamoto[1,2]*, Jiakun Li[1]*, Jeffrey A. O. Garber[1,2], Julian D. Rolfes[1,2], Gregory B. Boursalian[1,2], Jannik C. Borghs[2], Christophe Genicot[3], Jérôme Jacq[3], Maurice van Gastel[1], Frank Neese[1] & Tobias Ritter[1,2]

Aryl fluorides are widely used in the pharmaceutical and agrochemical industries[1,2], and recent advances have enabled their synthesis through the conversion of various functional groups. However, there is a lack of general methods for direct aromatic carbon–hydrogen (C–H) fluorination[3]. Conventional methods require the use of either strong fluorinating reagents, which are often unselective and difficult to handle, such as elemental fluorine, or less reactive reagents that attack only the most activated arenes, which reduces the substrate scope. A method for the direct fluorination of aromatic C–H bonds could facilitate access to fluorinated derivatives of functional molecules that would otherwise be difficult to produce. For example, drug candidates with improved properties, such as increased metabolic stability or better blood–brain-barrier penetration, may become available. Here we describe an approach to catalysis and the resulting development of an undirected, palladium-catalysed method for aromatic C–H fluorination using mild electrophilic fluorinating reagents. The reaction involves a mode of catalysis that is unusual in aromatic C–H functionalization because no organometallic intermediate is formed; instead, a reactive transition-metal-fluoride electrophile is generated catalytically for the fluorination of arenes that do not otherwise react with mild fluorinating reagents. The scope and functional-group tolerance of this reaction could provide access to functional fluorinated molecules in pharmaceutical and agrochemical development that would otherwise not be readily accessible.

Conventional methods for aromatic fluorination require elemental fluorine or similarly reactive reagents, which are unselective and require specialized equipment to handle safely[4]. Bench-stable electrophilic fluorinating reagents—such as *N*-fluoropyridinium salts, *N*-fluorobenzenesulfonimide (NFSI) and Selectfluor—are easier to handle but less reactive, and require either very electron-rich arenes or multiple equivalents of the arene to accomplish direct C–H fluorination[5,6]. Catalysis of aromatic C–H fluorination reactions has been reported using coordination-assistance to promote fluorination proximal to Lewis-basic functional groups, but such approaches are limited in scope to those substrates containing the required directing groups[7–10]. Advances in aliphatic C–H fluorination have been made[11], but currently there is no method for direct aromatic C–H fluorination with broad scope.

In our investigation into the catalysis of aromatic C–H fluorination reactions, we sought an approach that was distinct from the common C–H activation sequence in which C–H metalation precedes functionalization; with few exceptions[12–14], the conventional approach[15] requires multiple equivalents of the arene substrate to promote C–H metalation in the absence of a coordinating directing group. Instead, we sought to design catalysts with ancillary ligands that would favour the oxidation of the complex before any interaction with the substrate, giving rise to a reactive, high-valent metal-fluoride intermediate that is electrophilic at fluorine and capable of oxidative fluorine transfer to arenes. We designed the Pd(II) complex **1**, ligated simultaneously by a tridentate (terpyridine, terpy) and a bidentate (2-chloro-1,10-phenanthroline, 2-Cl-phen) ligand, which would be oxidized by electrophilic fluorinating reagents to yield the desired Pd(IV)–F complex **2** (Fig. 1b). The oxidation of doubly cationic **1** is promoted by a destabilizing interaction between the lone pair of the apical donor atom and the filled d_{z^2} orbital on Pd(II), which is readily apparent in the highest occupied molecular orbital of **1** as calculated by density functional theory (DFT) (Fig. 1c). X-ray diffraction corroborates the apical interaction: the

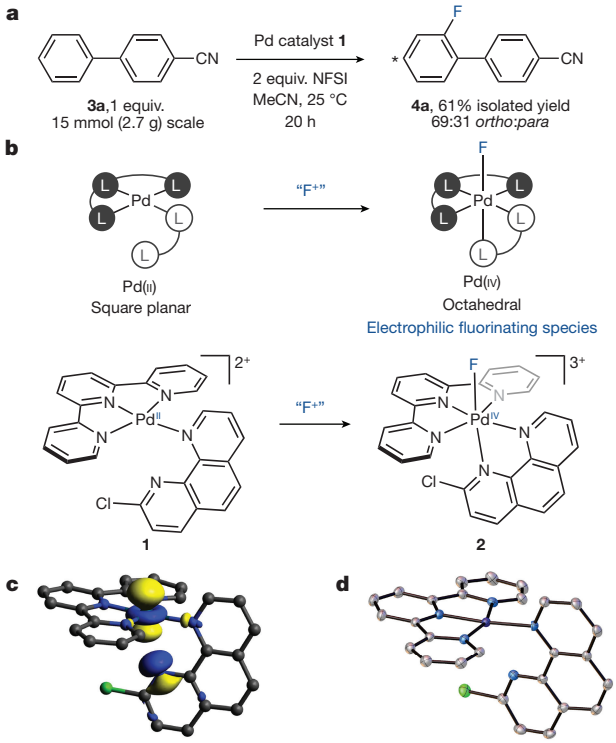


Figure 1 | Aromatic fluorination catalysed by 1. a, Palladium catalyst **1** enables the direct, non-chelation-assisted fluorination of 4-cyanobiphenyl. **b**, Oxidation of Pd(II) complex **1**, assisted by the ligand combination of terpy and 2-Cl-phen, yields the triply cationic Pd(IV)–F electrophile **2**. **c**, The highest occupied molecular orbital of **1**, as calculated by DFT, showing destabilizing orbital interaction. Hydrogen atoms are omitted for clarity. Calculations at the CPCM(MeCN)TPSS0 D3/def2-QZVP//PBE0 D3/def2-TZVP level of theory; iso = 0.05. **d**, X-ray crystal structure of **1**, shown with 50%-probability ellipsoids. Hydrogen atoms and solvent molecules are omitted for clarity.

[1]Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Mülheim an der Ruhr, Germany. [2]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA. [3]Global Chemistry, UCB NewMedicines, UCB Biopharma, 1420 Braine-L'Alleud, Belgium.
*These authors contributed equally to this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

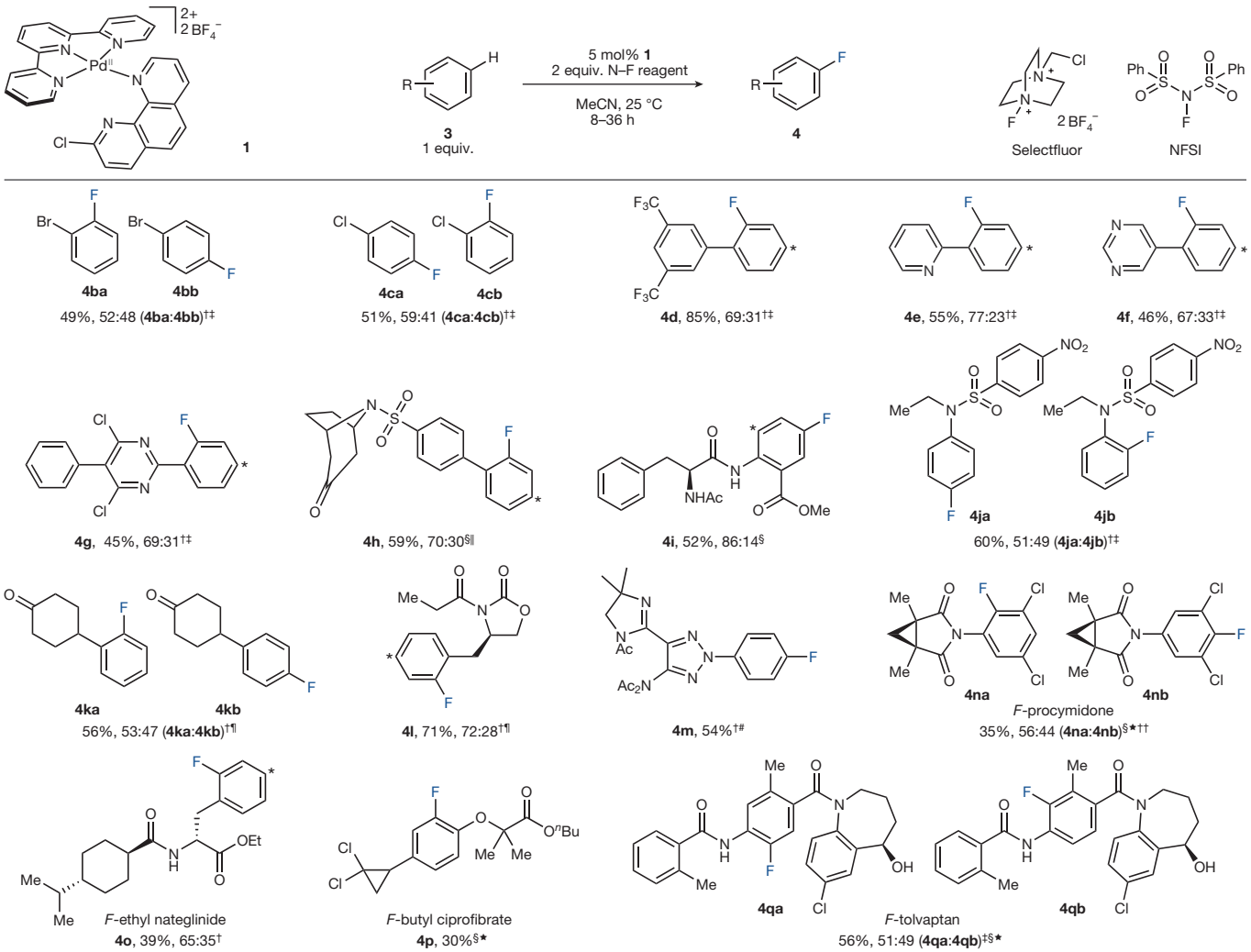


Figure 2 | Substrate scope of the Pd-catalysed fluorination of arenes. Reaction conditions: arene, 1 mmol; catalyst **1,** 5 mol%; Selectfluor or NFSI, 2 equiv.; MeCN, 0.1 M. All isomers of non-volatile products have been isolated and characterized as analytically pure samples. The asterisk denotes the site of fluorination of the constitutional isomer that is not shown. Overall yields and the ratio of the constitutional isomers are based on ^{19}F NMR integration of reaction mixtures with internal standard. †Reaction performed with Selectfluor. ‡Reaction performed at 50 °C. §Reaction performed with NFSI. ‖Reaction performed with 1,2-dichloroethane and MeCN (1:1, 0.1 M). ¶Reaction performed at 0 °C. #10 mol% catalyst **1** was used. ★7.5 mol% catalyst **1** was used. ††Reaction performed at 80 °C.

Pd–N distance in the X-ray structure of **1** (Fig. 1d) is 2.6 Å, which is 1.2 Å shorter than the sum of the van der Waals radii of Pd and N (ref. 16).

Complex **1** is a competent catalyst for the fluorination of various arenes by either Selectfluor or NFSI. The substrates shown in Fig. 2 underwent fluorination in the presence of **1** at room temperature or at 50 °C, but little background reactivity was observed in the absence of **1** under otherwise identical conditions (maximum <1% yield) or even under reflux in acetonitrile (maximum 21% yield). In our studies, catalyst **1** was formed *in situ* from $[Pd(terpy)(MeCN)](BF_4)_2$ and 2-Cl-phen, but it can also be generated by combining the commercially available palladium source $[Pd(MeCN)_4](BF_4)_2$ and the ligands before the addition of the reactants. Compatible functional groups include nitriles (**3a**), aryl bromides (**3b**), chlorides (**3c**, **3g, 3n, 3p, 3q**), certain heterocycles (**3e**–**3g**, **3m**), sulfonamides (**3h**, **3j**), ketones (**3h**, **3k**), amides (**3i**, **3l**–**3o, 3q**), esters (**3i**, **3o, 3p**), carbamates (**3l**), ethers (**3p**) and free hydroxy groups (**3q**). Five-membered heteroarenes containing nitrogen (**3m**) can be tolerated, but oxidatively labile functional groups such as amines and thiols cannot, owing to their general incompatibility with electrophilic fluorinating reagents. Electron-deficient arenes such as **3b**–**3d** are not successfully fluorinated through conventional methodologies, but are suitable substrates for fluorination via catalysis with **1**. However, more electron-deficient arenes, such as methyl benzoate, are insufficiently reactive and undergo little or no conversion.

Structurally complex substrates—such the pesticide procymidone (**3n**), the type-2 diabetes drug nateglinide (**3o**), the lipid-lowering agent ciprofibrate (**3p**) and the hyponatremia drug tolvaptan (**3q**)—were fluorinated directly via catalysis with **1**. Although fluorine can impart desirable properties on pharmaceuticals and agrochemicals, fluorinated analogues of structurally complex molecules can currently be difficult to access; conventional fluorination methods failed to provide the fluorinated products shown in Fig. 2.

In most cases, the fluorination reaction affords mixtures of at least two constitutional isomers, resulting from similar rates of fluorination at the positions *ortho* and *para* to the aromatic substituents. Purification of aryl fluoride products from mixtures of their constitutional isomers and the starting material is often challenging[17,18]. However, the isolation and characterization of all the non-volatile products obtained here has been achieved, although optimization of the separation protocol was required for each substrate. For example, the *ortho*- and *para*-fluorinated products of the gram-scale fluorination of 4-cyanobiphenyl have been separated in 61% isolated yield (Fig. 1a). Although high positional selectivity is generally desired in C–H functionalization reactions, mixtures of constitutional isomers can be advantageous for some applications. For example, in the late-stage derivatization of drug candidates, each product isomer is an additional derivative that can be obtained without the need for costly and laborious *de novo* synthesis[19,20]. Fluorinated tolvaptan

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

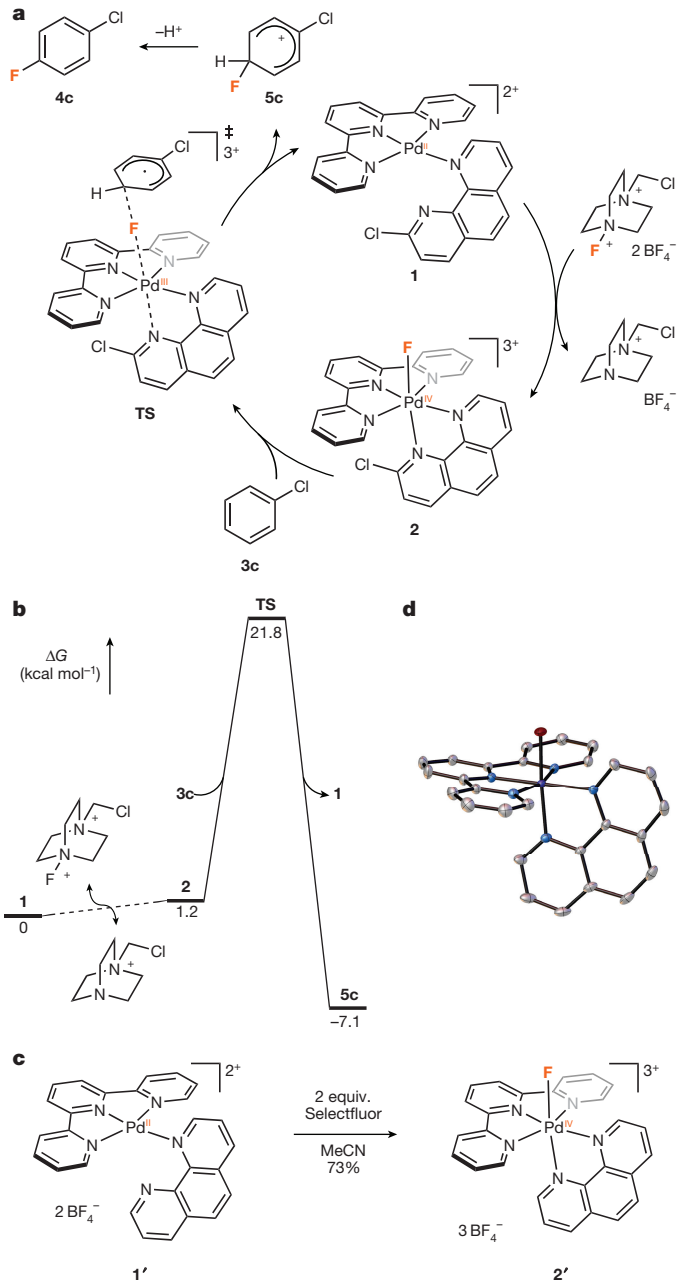


Figure 3 | Mechanism of fluorination catalysed by 1. a, The proposed catalytic cycle for the fluorination of chlorobenzene (**3c**). **b**, Energy-level diagram of the proposed catalytic cycle with chlorobenzene (**3c**); energies calculated by DFT. **c**, Synthesis of Pd(IV)–F complex **2′** via the oxidation of **1′** by Selectfluor. **d**, X-ray crystal structure of **2′**, shown with 50%-probability ellipsoids. Pd–F bond length: measured, 1.9120(7) Å; calculated, 1.89 Å.

(**4q**), for example, would be challenging and time-consuming to prepare with conventional chemistry through *de novo* syntheses. Late-stage fluorination, even with the requirement for a custom-made separation protocol, can conveniently produce promising new candidates that may have never been evaluated otherwise.

The proposed mode of action of **1** is highly unusual in the catalysis of aromatic oxidation reactions: conceptually, an activated electrophile is generated *in situ* from **1**, in the form of Pd(IV)–F intermediate **2**. The activated Pd(IV)–F electrophile **2** would therefore be capable of electrophilic fluorination of weakly nucleophilic arenes that cannot be fluorinated directly by Selectfluor and NFSI (Fig. 3a)[21,22]. A DFT analysis of the aryl fluorination reaction suggests a mechanism that is accessible to Pd(IV)–F **2** but not to Selectfluor or NFSI. On the basis of these results, we hypothesize that the fluorination mechanism proceeds through a single transition state via fluoride-coupled electron transfer (Fig. 3a, b).

Selectfluor fluorinates only electron-rich arenes such as anisole[23]; complex **1**, conversely, is able to catalyse the fluorination of electron-deficient arenes such as chlorobenzene. DFT calculations suggest that Pd(IV)–F **2** has a higher single-electron reduction potential than that of Selectfluor, although both compounds have a similar thermodynamic driving force for electrophilic fluorination. The transition state **TS** of the fluorination of chlorobenzene with Pd(IV)–F **2** shows high spin-density on the Pd as well as on the aryl carbon atoms. As such, the transition state is most appropriately characterized as a singlet diradical; two subsequent fluoride-coupled electron transfers occur asynchronously as the reaction proceeds through a single transition state. The mechanism is reminiscent of that previously reported for the fluorination of enamines and organometallic reagents with an isolated Pd(IV)–F (ref. 24). The calculated energy barrier for electrophilic fluorination of 21.8 kcal mol^{-1} (Fig. 3b) is in agreement with the observed reaction time and the temperature of the reaction (as discussed below, see Supplementary Information for details).

We sought to produce Pd(IV)–F complex **2**, to verify our design principle as well as to investigate the reactivity of **2** with arenes. Treatment of **1** with XeF_2 in the presence of $LiBF_4$ produced a ^{19}F NMR signal attributable to **2** at $\delta = -258.6$ p.p.m.; however, complex **2** is reduced to **1** in acetonitrile solution, even in the absence of substrate, which hindered our attempts at isolation. Treatment of **1** with Selectfluor over a range of temperatures (−40 °C to 25 °C) resulted in the desired reduction of Selectfluor, but did not produce a ^{19}F NMR signal that could be attributed to a Pd(IV)–F species, presumably because **2** is reduced faster than it is formed under these conditions, and therefore does not accumulate in observable quantities.

Complex **1′**, in which 2-Cl-phen is replaced with unsubstituted phenanthroline (phen), is a competent catalyst for aromatic fluorination, although not as effective as catalyst **1**. When **1′** was treated with Selectfluor in acetonitrile at room temperature and then allowed to stand at −35 °C, the Pd(IV)–F complex **2′** (^{19}F NMR −259.5 p.p.m.) precipitated in 73% yield (Fig. 3c). Complex **2′** is sufficiently stable at low temperature to enable characterization and reactivity studies. The higher stability of **2′** as compared to **2** is understood to result from the greater electron-donating ability of phen relative to that of 2-Cl-phen. Likewise, the greater reactivity of **2** may explain why the 2-Cl-phen-ligated complex **1** outperforms the phen-ligated **1′** in terms of catalytic ability.

Pd(IV)–F complex **2′** reacts with arenes to yield fluorinated products (Fig. 4). For example, 4-cyanobiphenyl (**3a**), when treated with **2′** in acetonitrile, yielded a 66:34 ratio of *ortho*- and *para*-fluoro isomers in 63% overall yield. The positional selectivity of fluorination by **2′** is similar to that observed in fluorination catalysed by **1′** (69:31 *ortho*:*para*), consistent with **2′** being the C–F-bond-forming species in catalysis by **1′**. These selectivity ratios are, in turn, similar to that observed upon fluorination with the optimal 2-Cl-phen-ligated catalyst **1** (71:29 *ortho*:*para*). We cannot rule out the involvement of a different C–F bond-forming pathway in catalysis by **1** (for example, through a Pd(III) species); however, the similar positional selectivities observed in fluorination catalysed by **1**, fluorination catalysed by **1′**, and stoichiometric fluorination by **2′** are consistent with fluorination by similar Pd(IV)–F species in all three cases, corresponding to **2** in the case of catalysis by **1**.

Aromatic C–H oxidations catalysed by transition metals generally proceed via C–H metalation followed by oxidation of the resulting organometallic intermediate, with product formation ensuing through reductive elimination. The aromatic fluorination catalysed by **1** presented here is an unusual example of an alternative mode of catalysis for aromatic C–H oxidation, in which a transition-metal catalyst is oxidized to a high-valent intermediate, which in turn oxidizes the substrate by group transfer of a ligand. Such an 'oxidation-first' mechanism is reminiscent of various transition-metal-catalysed aliphatic C–H oxidations, such as hydroxylations[25–27] and halogenations[28] through high-valent metal–oxo complexes, and aminations[29] via metal–nitrenoid

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

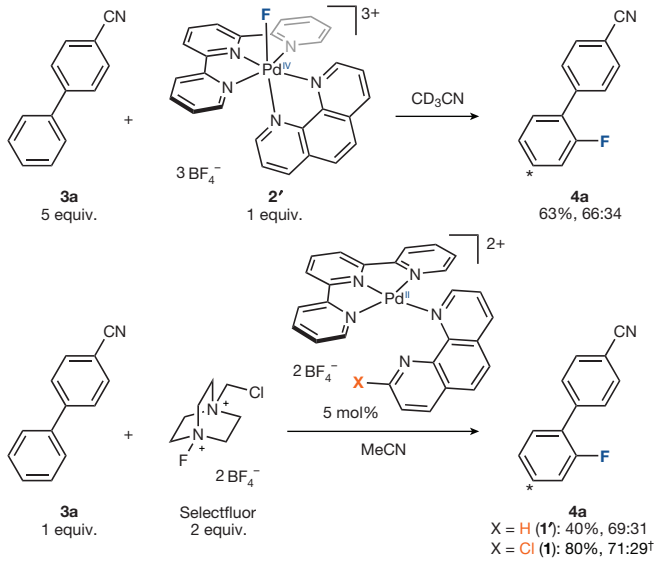


Figure 4 | Comparison of the positional selectivity of stoichiometric and catalytic fluorinations using 2′. Top, stoichiometric fluorination; bottom, catalytic fluorination. †5 equiv. of 4-cyanobiphenyl (**3a**) and 1 equiv. of Selectfluor were used.

species. Reported examples of such a mechanism for aromatic C–H oxidation, however, are rare, and are proposed to involve metal-nitrene or aminyl-radical transfer to the arene, although evidence for the proposed modes of action in these cases is indirect[30,31]. To the best of our knowledge, the aromatic fluorination reaction reported here is the only example so far of a synthetic method for aromatic C–H functionalization in which oxidation reactivity between a high-valent catalytic intermediate and arenes has been directly scrutinized.

The other examples of aromatic and aliphatic C–H oxidation reactions proceeding through 'oxidation-first' mechanisms mentioned above function because the high-valent intermediate provides access to a mechanism of oxidation that is not available to the starting reagents, such as a radical-rebound mechanism in the case of hydroxylation through metal–oxo intermediates. We have shown here data that support the interpretation that catalyst **2** provides access to a fluoride-coupled electron-transfer mechanism that is not accessible to electrophilic fluorinating reagents such as Selectfluor.

We anticipate that the direct electrophilic C–H fluorination of arenes reported here will be a useful tool in medicinal chemistry; indeed, the reaction has already found use in the late-stage derivatization of drug molecules. Furthermore, the unusual mechanism of catalysis by complex **1**, in which a high-valent transition-metal intermediate undergoes group transfer to arenes, may become the basis for a new approach to the catalysis of C–H functionalization reactions.

Data Availability Data are available from the corresponding author on reasonable request.

Received 21 March; accepted 17 December 2017.

1. Müller, K., Faeh, C. & Diederich, F. Fluorine in pharmaceuticals: looking beyond intuition. *Science* **317,** 1881–1886 (2007).
2. Gillis, E. P., Eastman, K. J., Hill, M. D., Donnelly, D. J. & Meanwell, N. A. Applications of fluorine in medicinal chemistry. *J. Med. Chem.* **58,** 8315–8359 (2015).
3. Campbell, M. G. & Ritter, T. Modern carbon–fluorine bond forming reactions for aryl fluoride synthesis. *Chem. Rev.* **115,** 612–633 (2015).
4. Sandford, G. Elemental fluorine in organic chemistry (1997–2006). *J. Fluor. Chem.* **128,** 90–104 (2007).
5. Taylor, S. D., Kotoris, C. C. & Hum, G. Recent advances in electrophilic fluorination. *Tetrahedron* **55,** 12431–12477 (1999).
6. Lal, G. S., Pez, G. P. & Syvret, R. G. Electrophilic NF fluorinating agents. *Chem. Rev.* **96,** 1737–1756 (1996).
7. Fier, P. S. & Hartwig, J. F. Selective C–H fluorination of pyridines and diazines inspired by a classic amination reaction. *Science* **342,** 956–960 (2013).
8. Chan, K. S. L., Wasa, M., Wang, X. & Yu, J.-Q. Palladium(II)-catalyzed selective monofluorination of benzoic acids using a practical auxiliary: a weak-coordination approach. *Angew. Chem. Int. Ed.* **50,** 9081–9084 (2011).
9. Wang, X., Mei, T.-S. & Yu, J.-Q. Versatile $Pd(OTf)_2{\cdot}2H_2O$-catalyzed *ortho*-fluorination using NMP as a promoter. *J. Am. Chem. Soc.* **131,** 7520–7521 (2009).
10. Hull, K. L., Anani, W. Q. & Sanford, M. S. Palladium-catalyzed fluorination of carbon–hydrogen bonds. *J. Am. Chem. Soc.* **128,** 7134–7135 (2006).
11. Liu, W. *et al.* Oxidative aliphatic C–H fluorination with fluoride ion catalyzed by a manganese porphyrin. *Science* **337,** 1322–1325 (2012).
12. Mkhalid, I. A. I., Barnard, J. H., Marder, T. B., Murphy, J. M. & Hartwig, J. F. C–H activation for the construction of C–B bonds. *Chem. Rev.* **110,** 890–931 (2010).
13. Cheng, C. & Hartwig, J. F. Rhodium-catalyzed intermolecular C–H silylation of arenes with high steric regiocontrol. *Science* **343,** 853–857 (2014).
14. Wang, P. *et al.* Ligand-accelerated non-directed C–H functionalization of arenes. *Nature* **551,** 489–493 (2017).
15. Kuhl, N., Hopkinson, M. N., Wencel-Delord, J. & Glorius, F. Beyond directing groups: transition-metal-catalyzed C–H activation of simple arenes. *Angew. Chem. Int. Ed.* **51,** 10236–10254 (2012).
16. Alvarez, S. A cartography of the van der Waals territories. *Dalton Trans.* **42,** 8617–8636 (2013).
17. Regalado, E. L., Makarov, A. A., McClain, R., Przybyciel, M. & Welch, C. J. Search for improved fluorinated stationary phases for separation of fluorine-containing pharmaceuticals from their desfluoro analogs. *J. Chromatogr. A* **1380,** 45–54 (2015).
18. Regalado, E. L. *et al.* Support of academic synthetic chemistry using separation technologies from the pharmaceutical industry. *Org. Biomol. Chem.* **12,** 2161–2166 (2014).
19. Hyohdoh, I. *et al.* Fluorine scanning by nonselective fluorination: enhancing Raf/MEK inhibition while keeping physicochemical properties. *ACS Med. Chem. Lett.* **4,** 1059–1063 (2013).
20. Cernak, T., Dykstra, K. D., Tyagarajan, S., Vachal, P. & Krska, S. W. The medicinal chemist's toolbox for late stage functionalization of drug-like molecules. *Chem. Soc. Rev.* **45,** 546–576 (2016); erratum **46,** 1760 (2017).
21. McCall, A. S. & Kraft, S. Pyridine-assisted chlorinations and oxidations by palladium(IV). *Organometallics* **31,** 3527–3538 (2012).
22. McCall, A. S., Wang, H., Desper, J. M. & Kraft, S. Bis-*N*-heterocyclic carbene palladium(IV) tetrachloride complexes: synthesis, reactivity, and mechanisms of direct chlorinations and oxidations of organic substrates. *J. Am. Chem. Soc.* **133,** 1832–1848 (2011).
23. Geng, C., Du, L., Liu, F., Zhu, R. & Liu, C. Theoretical study on the mechanism of selective fluorination of aromatic compounds with Selectfluor. *RSC Adv.* **5,** 33385–33391 (2015).
24. Brandt, J. R., Lee, E., Boursalian, G. B. & Ritter, T. Mechanism of electrophilic fluorination with Pd(IV): fluoride capture and subsequent oxidative fluoride transfer. *Chem. Sci.* **5,** 169–179 (2014).
25. Groves, J. T. High-valent iron in chemical and biological oxidations. *J. Inorg. Biochem.* **100,** 434–447 (2006).
26. McNeill, E. & Du Bois, J. Ruthenium-catalyzed hydroxylation of unactivated tertiary C–H bonds. *J. Am. Chem. Soc.* **132,** 10202–10204 (2010).
27. Chen, M. S. & White, M. C. A predictably selective aliphatic C–H oxidation reaction for complex molecule synthesis. *Science* **318,** 783–787 (2007).
28. Liu, W. & Groves, J. T. Manganese catalyzed C–H halogenation. *Acc. Chem. Res.* **48,** 1727–1735 (2015).
29. Du Bois, J. Rhodium-catalyzed C–H amination. An enabling method for chemical synthesis. *Org. Process Res. Dev.* **15,** 758–762 (2011).
30. Boursalian, G. B., Ngai, M.-Y., Hojczyk, K. N. & Ritter, T. Pd-catalyzed aryl C–H imidation with arene as the limiting reagent. *J. Am. Chem. Soc.* **135,** 13278–13281 (2013).
31. Paudyal, M. P. *et al.* Dirhodium-catalyzed C–H arene amination using hydroxylamines. *Science* **353,** 1144–1147 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Gitlin for HPLC purification, R. Goddard, S. Palm and J. Rust for X-ray crystallographic analysis, C. Farès for NMR spectroscopy, and M. Blumenthal, D. Kampen and S. Marcus for mass spectrometry. We thank UCB Biopharma for funding and compound separation, the Japan Society for the Promotion of Science and L'Oréal-UNESCO Japan for graduate fellowships to K.Y., the Fonds der Chemischen Industrie for a graduate fellowship for J.D.R., and the German Academic Exchange Service, DAAD for an Otto-Bayer fellowship to J.C.B.

Author Contributions K.Y. designed catalyst **1** and optimized the fluorination reaction. K.Y. and J.D.R. conducted mechanistic studies. K.Y. and G.B.B. developed the conceptual approach to the project. J.L., J.A.O.G., J.C.B., K.Y. and J.D.R. explored the substrate scope. J.D.R. performed DFT calculations with input from M.v.G. and F.N. G.B.B. wrote the manuscript with input from all other authors. C.G. and J.J. supported development towards useful examples. K.Y., J.L., J.A.O.G., G.B.B., J.D.R. and T.R. analysed the data. T.R. directed the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to T.R. (ritter@kofo.mpg.de).

Reviewer Information *Nature* thanks J. Groves and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature25493

Meridional overturning circulation conveys fast acidification to the deep Atlantic Ocean

Fiz F. Perez[1]*, Marcos Fontela[1]*, Maribel I. García-Ibáñez[1]*, Herlé Mercier[2]*, Anton Velo[1], Pascale Lherminier[2], Patricia Zunino[2], Mercedes de la Paz[1], Fernando Alonso-Pérez[1], Elisa F. Guallart[1] & Xose A. Padin[1]

Since the Industrial Revolution, the North Atlantic Ocean has been accumulating anthropogenic carbon dioxide (CO_2) and experiencing ocean acidification[1], that is, an increase in the concentration of hydrogen ions (a reduction in pH) and a reduction in the concentration of carbonate ions. The latter causes the 'aragonite saturation horizon'—below which waters are undersaturated with respect to a particular calcium carbonate, aragonite—to move to shallower depths (to shoal), exposing corals to corrosive waters[2,3]. Here we use a database analysis to show that the present rate of supply of acidified waters to the deep Atlantic could cause the aragonite saturation horizon to shoal by 1,000–1,700 metres in the subpolar North Atlantic within the next three decades. We find that, during 1991–2016, a decrease in the concentration of carbonate ions in the Irminger Sea caused the aragonite saturation horizon to shoal by about 10–15 metres per year, and the volume of aragonite-saturated waters to reduce concomitantly. Our determination of the transport of the excess of carbonate over aragonite saturation ($_{xc}[CO_3^{2-}]$)—an indicator of the availability of aragonite to organisms—by the Atlantic meridional overturning circulation shows that the present-day transport of carbonate ions towards the deep ocean is about 44 per cent lower than it was in preindustrial times. We infer that a doubling of atmospheric anthropogenic CO_2 levels—which could occur within three decades according to a 'business-as-usual scenario' for climate change[4]—could reduce the transport of $_{xc}[CO_3^{2-}]$ by 64–79 per cent of that in preindustrial times, which could severely endanger cold-water coral habitats. The Atlantic meridional overturning circulation would also export this acidified deep water southwards, spreading corrosive waters to the world ocean.

Atmospheric CO_2 levels have increased from 280 to 400 parts per million (p.p.m.) since the Industrial Revolution. The global ocean has captured some 30% of this anthropogenic CO_2 (C_{ant}), thereby acting as a climate regulator[1]. This CO_2 absorption has led to a decrease in seawater pH (of about 0.12 units) and in the supersaturation of calcium carbonate ($CaCO_3$)—effects known collectively as ocean acidification. Ocean acidification can particularly affect marine calcifiers[1] by favouring the dissolution of $CaCO_3$-based shells and skeletons[2]. Notably, deep cold-water coral (CWC) reefs formed by scleractinian corals—such as the ecosystem engineer *Lophelia pertusa*—are highly vulnerable to ocean acidification[3]. The global distribution of CWC seems to be partly limited by the depth of the aragonite saturation horizon (ASH), which marks the boundary between aragonite-stable waters above and dissolution-prone waters below[5]. In preindustrial times, more than 95% of CWC locations were found above the ASH[6], providing evidence that environments located below the ASH are hostile to CWC growth. In the North Pacific Ocean (Fig. 1a), where the ASH is only 500 metres deep[7], the distribution of aragonitic CWC is patchy and CWCs do not develop to form the large, deep reef frameworks that are abundant in the North Atlantic[8], where the ASH has tended to occur at depths of more than 2,000 metres. However, ocean acidification is causing the ASH to shoal, thus exposing CWCs to $CaCO_3$ undersaturation. Although laboratory experiments suggest that adult *L. pertusa* can acclimatize to $CaCO_3$ undersaturation[9,10], the long-term survival of CWC reefs in undersaturated water is questioned, because *L. pertusa* skeletons become weaker when exposed yearlong to levels of ocean acidification that are predicted to occur in the future[10], and the dead skeletal framework that supports the reef itself is likely to dissolve in undersaturated waters[11].

To determine the degree of aragonite saturation of the world ocean waters, we calculated $_{xc}[CO_3^{2-}]$ (in $\mu mol\,kg^{-1}$) by using quality-controlled global data sets of marine CO_2 system measurements[12,13]. A positive (or negative) $_{xc}[CO_3^{2-}]$ indicates waters that are supersaturated (or undersaturated) with aragonite (see Methods). We find that high positive $_{xc}[CO_3^{2-}]$ values occur in the North Atlantic while negative values occur in the North Pacific; this is consistent with the known distribution of CWC[14] below 1,000 metres (Fig. 1a). About 61% of CWCs found below 1,000 metres are located in the North Atlantic (this number rises to 78% at depths below 1,500 metres), where $_{xc}[CO_3^{2-}]$ is greater than 0, with an average $_{xc}[CO_3^{2-}]$ of $24.5\,\mu mol\,kg^{-1}$ below 1,000 metres (and $15\,\mu mol\,kg^{-1}$ below 1,500 metres). These are half the natural (preindustrial) $_{xc}[CO_3^{2-}]$ values (see Extended Data Table 1). The Atlantic meridional overturning circulation (AMOC) created the favourable conditions for CWC growth in the North Atlantic by conveying ventilated waters loaded with relatively high pH and positive $_{xc}[CO_3^{2-}]$ to the deep Atlantic Ocean.

In the centre of the Irminger Sea (Fig. 1b), winter deep convection recorded during 1991–2016 explains a persistent increase in C_{ant} from 30 to $50\,\mu mol\,kg^{-1}$ and the associated deep injection of ocean acidification in the ventilated subpolar mode water (Fig. 2a). During the same period, the atmospheric C_{ant} grew from 85 p.p.m. to 123 p.p.m. A thick layer of low salinity (less than 34.91) traces the strong convection events that occurred during the first half of the 1990s and during 2014–2016. During those strong convection events, subpolar mode water was ventilated down to 1,500 metres, showing high temporal variability and no indication of a slowing-down of deep convection[15,16]. The present-day surface ocean has about a 30% higher concentration of H^+ ions ($[H^+]$) than the natural (preindustrial) surface ocean[1]. At the centre of the Irminger Sea, the $[H^+]$ increase affects a layer about 1,500 metres thick (Fig. 2c)—the deepest signal yet seen of the direct injection of ocean acidification. Since 2002, anthropogenic perturbations have caused the isolines that mark 25% and 30% anthropogenic $[H^+]$ to deepen from the surface down to 1,500 metres. The isolines that mark $_{xc}[CO_3^{2-}]$ progressively ascend at about 10–15 metres per year (Fig. 2b), with some slightly faster ascension periods that are related to deep convection events (arrows in Fig. 2b). The effect of these deep convection events is buffered by

[1]Instituto Investigaciones Marinas (IIM, CSIC), calle Eduardo Cabello, 6, 36208, Vigo, Spain. [2]Centre National de la Recherche Scientifique (CNRS), Ifremer, Université de Brest, Institut de Recherche pour le Développement, Laboratoire d'Océanographie Physique et Spatiale (LOPS), Centre Ifremer de Bretagne, 29280, Plouzané, France.
*These authors contributed equally to this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

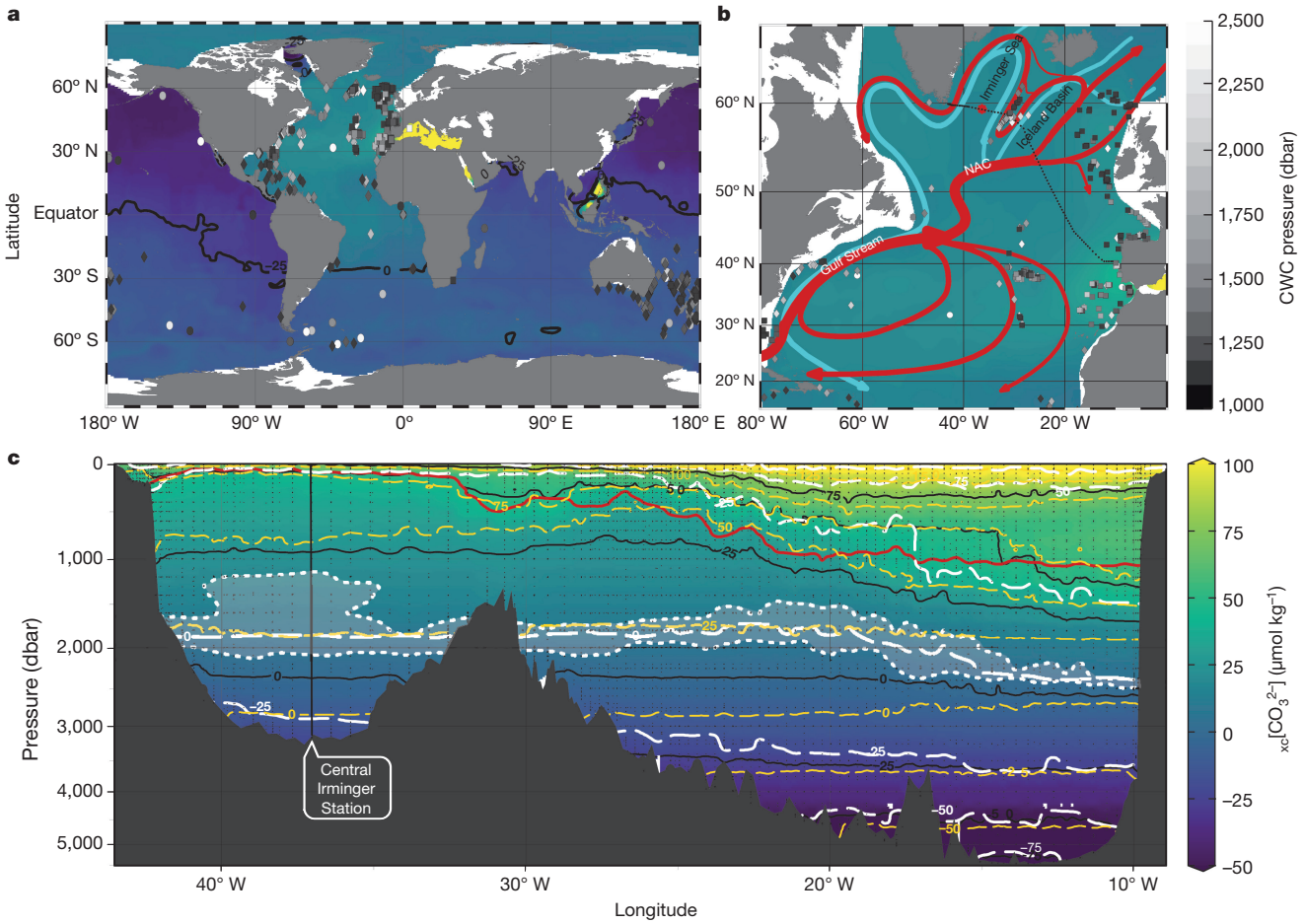


Figure 1 | Aragonite saturation and cold-water coral distribution. **a**, $_{xc}[CO_3^{2-}]$ at 1,500 metres (see the colour bar at the bottom right). Circles, diamonds and squares represent CWC locations (below 1,000 m) where $_{xc}[CO_3^{2-}]$ is, respectively, less than 0 μmol kg^{-1}, between 0 μmol kg^{-1} and 25 μmol kg^{-1}, or more than 25 μmol kg^{-1}. Depths are coded in grey (see the colour bar at the top right). **b**, Map showing the Central Irminger Station (red star), the Ovide section (dotted line), and the circulation of upper (red) and lower (blue) AMOC limbs. **c**, Along the Ovide section, $_{xc}[CO_3^{2-}]$ observed (black lines, and coloured shading as indicated in the colour bar to the right), natural (preindustrial; yellow dashed lines), and computed for a doubling of present atmospheric C_{ant} (white dashed line). The grey area shows the uncertainties in the ASH (where $_{xc}[CO_3^{2-}] = 0$). The red line separates the upper and lower AMOC limbs. Dots represent samples. The dark shaded area represents the sea bed. Figure prepared using Ocean Data View/DIVA.

the partial balance between the increase in anthropogenic [H$^+$] and the decrease in natural [H$^+$] (the latter resulting from the removal of the natural CO_2 accumulated in old waters by the remineralization of organic matter) (Fig. 2c). Because of this biogeochemical feedback, the decrease in $_{xc}[CO_3^{2-}]$ and the associated shoaling of the ASH are only weakly sensitive to the intensity of the deep convection.

West of 20° W in the subpolar North Atlantic (SPNA), the ASH observed in 2016 is 200–500 metres shallower than the preindustrial ASH (Fig. 1c). Along the 2016 Greenland–Portugal Ovide transect[17] (Fig. 1b), the 25 μmol kg^{-1} isoline of the present-day $_{xc}[CO_3^{2-}]$ runs close to the 50 μmol kg^{-1} isoline of natural $_{xc}[CO_3^{2-}]$, indicating a 50% decrease in the availability of CO_3^{2-} ions for deep CWC. These changes reflect the spreading of newly ventilated subpolar mode water with low $_{xc}[CO_3^{2-}]$ from the Irminger Sea to the Iceland basin.

During 2002–2016, the mean transport of sea water by the AMOC across the Ovide section[17] was 16.2 ± 0.8 Sverdrup (Sv, where 1 Sv equals 1×10^6 m^3 s^{-1}). The upper and lower limbs of the AMOC showed contrasting $_{xc}[CO_3^{2-}]$ (Fig. 1c) and C_{ant} values (Extended Data Fig. 1), with average $_{xc}[CO_3^{2-}]$ values of 61 ± 2 μmol kg^{-1} and 28 ± 3 μmol kg^{-1} for the upper and lower limbs, respectively. The upper limb of the AMOC flows northwards (above the isopycnal that marks a water density of 1,032.15 kg m^{-3}), transporting 1,064 kmol s^{-1} of $_{xc}[CO_3^{2-}]$ (Fig. 3). About 40% of this transport (435 kmol s^{-1}) subducted into the lower limb of the AMOC in the SPNA, while the remaining 60% was exported to the Nordic Seas[18,19]. These $_{xc}[CO_3^{2-}]$ transport values are between 33% and 44% lower than the transport of natural $_{xc}[CO_3^{2-}]$, because natural $_{xc}[CO_3^{2-}]$ values in the upper and lower limbs of the AMOC (90 ± 2 and 49 ± 4 μmol kg^{-1}, respectively) are much larger than the present-day values (61 ± 2 and 28 ± 3 μmol kg^{-1}, respectively). By vertically transporting lower $_{xc}[CO_3^{2-}]$ with respect to preindustrial levels, the AMOC is responsible for the decreases in pH and in $_{xc}[CO_3^{2-}]$ in the deep North Atlantic.

Projections made by the Intergovernmental Panel on Climate Change (IPCC)[20] indicate that a warming of about 2 °C could be reached when atmospheric CO_2 levels rise to 480–520 p.p.m. (Extended Data Fig. 2)—that is, when there is a CO_2 excess over preindustrial values (C_{ant}) of 200–240 p.p.m. These figures are around double the present-day anthropogenic perturbations in temperature and CO_2 (1 °C warming[21] and a C_{ant} of 100–120 p.p.m.). We inferred the $_{xc}[CO_3^{2-}]$ that would result from a doubling of the atmospheric C_{ant} by adding 0.72 times the present-day marine C_{ant} (Extended Data Fig. 1) to the marine [CO_2] observed in 2002–2016, and assuming that C_{ant} changes follow a transient steady approximation[22,23] (see Methods). The projected slow-down of deep convection by the end of this century should not invalidate this approach because, as concluded above, the decrease in $_{xc}[CO_3^{2-}]$ is only weakly sensitive to the intensity of the deep convection. We find that, when doubling the atmospheric C_{ant}, the volume of aragonite-saturated deep waters in the Iceland basin and the Irminger

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

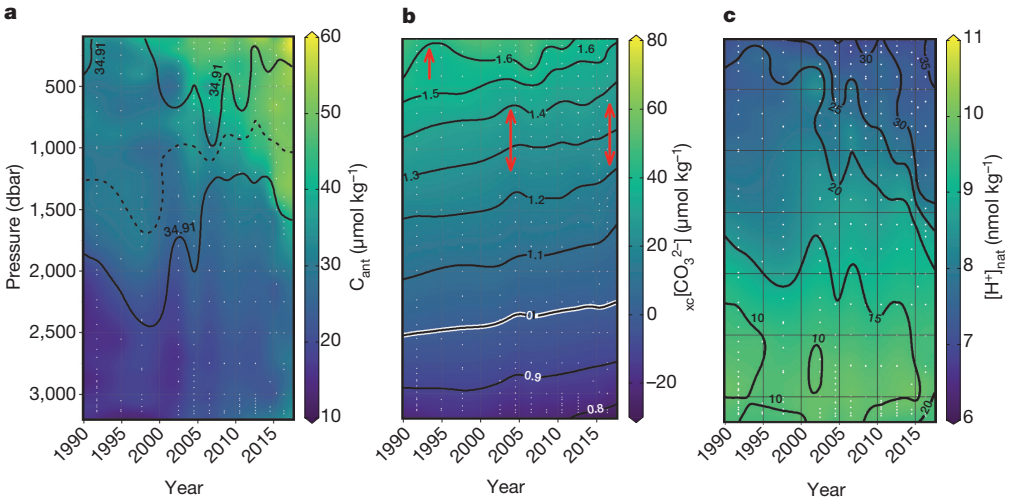


Figure 2 | Ocean acidification at the Central Irminger Station. a, Time evolution from 1991 to 2016 of: anthropogenic CO_2 (C_{ant}, colour bar); the isohaline of 34.91 (continuous black lines); and the salinity minimum (dotted black line). **b**, Time evolution from 1991 to 2016 of: aragonite saturation (contours); and the excess of CO_3^{2-} concentration over aragonite saturation ($_{xc}[CO_3^{2-}]$, colour bar). Red arrows indicate the effects of deep convection events. **c**, Time evolution from 1991 to 2016 of: natural (pre-industrial) H^+ concentration ($[H^+]_{nat}$, colour bar); and the percentage change in $[H^+]$ that results from anthropogenic emissions (contours). Dots represent samples. Figure prepared using Ocean Data View/DIVA.

Sea decreases dramatically, as shown by the fact that both the ASH and the isolines that mark 25 µmol kg^{-1} and 50 µmol kg^{-1} $_{xc}[CO_3^{2-}]$ shoal at 1,000–1,700 metres with respect to preindustrial depths (Fig. 1c). A striking feature is the disappearance of the vertical gradient of $[H^+]$ as a result of ocean acidification (Extended Data Fig. 3); this disappearance occurs because the relatively rapid deep injection of C_{ant} by the AMOC leads to the homogenization of the pH in the whole water column. This would in turn lead to a substantial weakening in the vertical gradient of $_{xc}[CO_3^{2-}]$.

We find that doubling the atmospheric C_{ant} could result in 70% of the CWC in the North Atlantic that live at depths lower than 1,500 metres finding themselves below the ASH, and thus exposed to negative $_{xc}[CO_3^{2-}]$ (Extended Data Table 1). An average reduction of 75% in $_{xc}[CO_3^{2-}]$ with respect to preindustrial levels would be observed for the CWC living below 1,000 metres. This might occur as soon as 2050 in the 'fossil-fuelled development' shared socioeconomic scenario (SSP5, which is comparable to the previous business-as-usual scenario (RCP8.5); see Extended Data Fig. 2)[4]. Our database results are in line with IPCC projections[24] that point to ocean acidification being a serious threat for deep (below 1,000 metre) CWC habitats in the North Atlantic by 2100. However, most recent model studies have shown large discrepancies in their estimates of the extent of CWC habitats that could be exposed to corrosive waters by the end of the century. The proportion of CWC projected to live below the ASH varies between 23% and 70% of its present-day area in the most recent simulations, using the business-as-usual scenario[6,25]. In another study[26], the proportion was projected to reach 85% of the Northeast Atlantic CWC surfaces by 2060. Our findings support this latter prediction. We show that the water masses of the lower limb of the AMOC are becoming more acidic. Those water masses are ventilating the deep layers of the

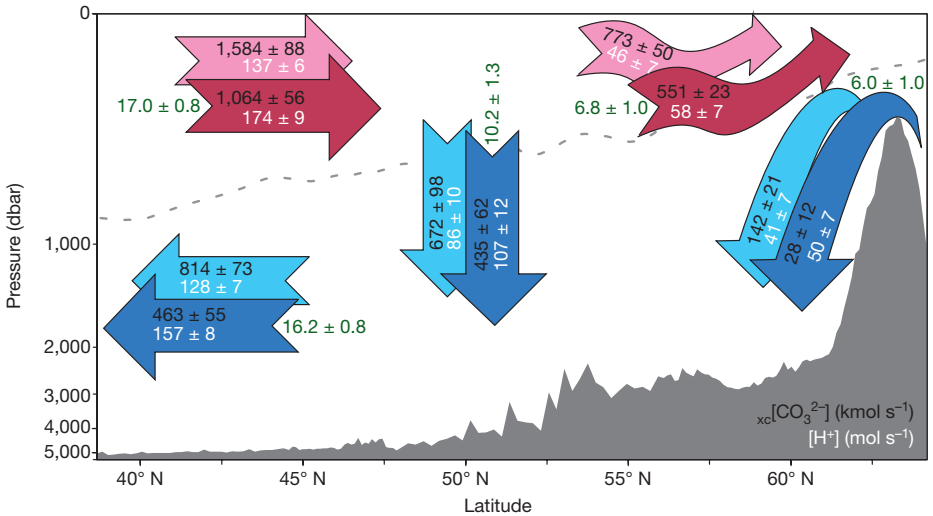


Figure 3 | Circulation of the excess of $[CO_3^{2-}]$ over aragonite saturation and of $[H^+]$ in the subpolar North Atlantic. Mean transports for 2002–2016 between the Ovide transect and the Nordic Sills of: seawater (in Sv; green values); natural (light coloured arrows) and observed (dark coloured arrows) excess of $[CO_3^{2-}]$ over aragonite saturation ($_{xc}[CO_3^{2-}]$, in kmol s^{-1}; black values); and natural (light coloured arrows) and observed (dark coloured arrows) $[H^+]$ (in mol s^{-1}; white values). Uncertainties are the errors in the mean transports across the eight occupations of the Ovide line during 2002–2016. The grey dotted line represents the limit between the upper and lower limbs of the AMOC. The dark shaded area represents the sea bed.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

world ocean and propagating the acidification threat to the CWC habitats of the world. We estimate that, if the atmospheric C_{ant} doubles, the average $_{xc}[CO_3^{2-}]$ below 1,000 metres would become negative globally (Extended Data Table 1), indicating that most of the deep CWC habitats of the world ocean might be exposed to corrosive waters by 2050.

Models predict that the mid-latitude AMOC could slow by 25% by the end of this century[20]; this would lead directly to a decreasing volume of waters with negative $_{xc}[CO_3^{2-}]$ being exported southwards to the global deep ocean. However, such an AMOC reduction would also lead to a reduction in the ventilation of the deep ocean at mid-latitudes and thus, through remineralization processes, to a reduction in both the dissolved oxygen and pH through remineralization processes[27–29]. Therefore, future AMOC slow-down could worsen the acidification threats for deep CWC.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 August 2017; accepted 3 January 2018.
Published online 12 February 2018.

1. Gattuso, J.-P. *et al.* Contrasting futures for ocean and society from different anthropogenic CO_2 emissions scenarios. *Science* **349,** aac4722 (2015).
2. Pörtner, H.-O. *et al.* in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change* (eds Field, C. B. *et al.*) 411–484 (Cambridge Univ. Press, 2014).
3. Roberts, J. M. *Cold-Water Corals: The Biology and Geology of Deep-Sea Coral Habitats* (Cambridge Univ. Press, 2009).
4. Riahi, K. *et al.* The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Glob. Environ. Change* **42,** 153–168 (2017).
5. Tittensor, D. P., Baco, A. R., Hall-Spencer, J. M., Orr, J. C. & Rogers, A. D. Seamounts as refugia from ocean acidification for cold-water stony corals. *Mar. Ecol.* **31,** 212–225 (2010).
6. Guinotte, J. M. *et al.* Will human-induced changes in seawater chemistry alter the distribution of deep-sea scleractinian corals? *Front. Ecol. Environ.* **4,** 141–146 (2006).
7. Jiang, L.-Q. *et al.* Climatological distribution of aragonite saturation state in the global oceans. *Glob. Biogeochem. Cycles* **29,** 1656–1673 (2015).
8. Feely, R. A. *et al.* Impact of anthropogenic CO_2 on the $CaCO_3$ system in the oceans. *Science* **305,** 362–366 (2004).
9. Maier, C., Hegeman, J., Weinbauer, M. G. & Gattuso, J.-P. Calcification of the cold-water coral *Lophelia pertusa*, under ambient and reduced pH. *Biogeosciences* **6,** 1671–1680 (2009).
10. Hennige, S. J. *et al.* Hidden impacts of ocean acidification to live and dead coral framework. *Proc. R. Soc. Lond. B* **282,** 20150990 (2015).
11. Thresher, R., Tilbrook, B., Fallon, S., Wilson, N. & Adkins, J. Effects of chronic low carbonate saturation levels on the distribution, growth and skeletal chemistry of deep-sea corals and other seamount megabenthos. *Mar. Ecol. Prog. Ser.* **442,** 87–99 (2011).
12. Olsen, A. *et al.* The global ocean data analysis project version 2 (GLODAPv2)—an internally consistent data product for the world ocean. *Earth Syst. Sci. Data* **8,** 297–323 (2016).
13. Key, R.M. *et al. Global Ocean Data Analysis Project, Version 2 (GLODAPv2)* https://www.nodc.noaa.gov/archive/arc0107/0162565/1.1/data/0-data/NDP_093.pdf (2015).
14. Freiwald, A., Fosså, H., Grehan, A., Koslow, T. & Roberts, J. M. *Cold-Water Coral Reefs: Out of Sight—No Longer Out of Mind* (UNEP-WCMC Biodiversity Series 22, Cambridge, 2004).
15. Pickart, R. S. *et al.* in *Arctic–Subarctic Ocean Fluxes* (eds Dickson, R. R., Meincke, J. & Rhines, P.) 629–652 (Springer, 2008).
16. Yashayaev, I. & Loder, J. W. Further intensification of deep convection in the Labrador Sea in 2016. *Geophys. Res. Lett.* **44,** 1429–1438 (2017).
17. Mercier, H. *et al.* Variability of the meridional overturning circulation at the Greenland–Portugal OVIDE section from 1993 to 2010. *Prog. Oceanogr.* **132,** 250–261 (2015).
18. Jeansson, E. *et al.* The Nordic Seas carbon budget: sources, sinks, and uncertainties. *Glob. Biogeochem. Cycles* **25,** GB4010 (2011).
19. Pérez, F. F. *et al.* Atlantic Ocean CO_2 uptake reduced by weakening of the meridional overturning circulation. *Nat. Geosci.* **6,** 146–152 (2013).
20. IPCC Working Group. *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2014).
21. Le Quéré, C. The implications of COP21 for our future climate. *Public Health Rev.* **37,** 29 (2016).
22. Tanhua, T. *et al.* Changes of anthropogenic CO_2 and CFCs in the North Atlantic between 1981 and 2004. *Glob. Biogeochem. Cycles* **20,** GB4017 (2006).
23. Steinfeldt, R., Rhein, M., Bullister, J. L. & Tanhua, T. Inventory changes in anthropogenic carbon from 1997–2003 in the Atlantic Ocean between 20° S and 65° N. *Glob. Biogeochem. Cycles* **23,** GB3010 (2009).
24. Hoegh-Guldberg, O. *et al.* in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change* (eds Barros, V. R. *et al.*) 1655–1731 (Cambridge Univ. Press, 2014).
25. Gehlen, M. *et al.* Projected pH reductions by 2100 might put deep North Atlantic biodiversity at risk. *Biogeosciences* **11,** 6955–6967 (2014).
26. Jackson, E. L., Davies, A. J., Howell, K. L., Kershaw, P. J. & Hall-Spencer, J. M. Future-proofing marine protected area networks for cold water coral reefs. *ICES J. Mar. Sci.* **71,** 2621–2629 (2014).
27. Tagklis, F., Bracco, A. & Ito, T. Physically driven patchy O_2 changes in the North Atlantic Ocean simulated by the CMIP5 Earth system models: CMIP5: modeled O_2 trends in the NA. *Glob. Biogeochem. Cycles.* **731,** 10.1002/2016GB005617 (2017).
28. Resplandy, L., Bopp, L., Orr, J. C. & Dunne, J. P. Role of mode and intermediate waters in future ocean acidification: analysis of CMIP5 models. *Geophys. Res. Lett.* **40,** 3091–3095 (2013).
29. Fontela, M., García-Ibáñez, M. I., Hansell, D. A., Mercier, H. & Pérez, F. F. Dissolved organic carbon in the North Atlantic meridional overturning circulation. *Sci. Rep.* **6,** 26931 (2016).

Acknowledgements The OVIDE research project was co-funded by the Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER) and CNRS/Institut National des Sciences de l'Univers (INSU)/Les Enveloppes Fluides et l'Environnement (LEFE). H.M. was supported by CNRS. This is a contribution to the AtlantOS project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement 633211. This study is also a contribution to the BOCATS project (CTM2013-41048-P) supported by the Spanish Ministry of Economy and Competitiveness and co-funded by the Fondo Europeo de Desarrollo Regional 2014–2020 (FEDER). We thank the captain of the research vessel *Sarmiento Gamboa*—A. Campos—and her crew for help that made possible the success of the BOCATS cruise. We thank D. Barton for revision of the manuscript.

Author Contributions F.F.P., H.M. and P.L. designed the program and executed the fieldwork. F.F.P, M.F., M.I.G-I., A.V., M.P., F.A-P., E.F.G. and X.A.P. contributed to the chemical determination of nutrients and carbon system. H.M., P.L. and P.Z. contributed to the measurement of currents and the inverse-model results. All authors contributed to discussion and writing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to F.F.P. (fiz.perez@iim.csic.es).

Reviewer Information *Nature* thanks J. Dunne, M. Roberts and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Aragonite saturation and carbonate ion excess. Aragonite is one of the metastable forms of $CaCO_3$ that is used by calcifying marine organisms to build their shells[30,31]. A measure of the biological availability of aragonite is the *in situ* degree of aragonite saturation (Ω_A), given by:

$$\Omega_A = \frac{[Ca^{2+}][CO_3^{2-}]_{is}}{K_A}$$

where $[Ca^{2+}]$ is the concentration of calcium ions; $[CO_3^{2-}]$ is the concentration of the carbonate ion; and K_A is the $CaCO_3$ aragonite solubility product. The subscript 'is' denotes *in situ* temperature and pressure conditions. $[Ca^{2+}]$ is conservative and is obtained from the salinity (S) as follows[32]:

$$[Ca^{2+}] = 0.01028 \times S/35$$

Another convenient measure of the aragonite saturation is the difference between $[CO_3^{2-}]_{is}$ and the $[CO_3^{2-}]$ at saturation ($[CO_3^{2-}]_{sat(\Omega_A=1)}$), which is the excess of carbonate ion concentration over aragonite saturation ($_{xc}[CO_3^{2-}]$):

$$_{xc}[CO_3^{2-}] = [CO_3^{2-}]_{is} - [CO_3^{2-}]_{sat(\Omega_A=1)}$$

A positive (or negative) $_{xc}[CO_3^{2-}]$ indicates that water is supersaturated (or undersaturated) with respect to aragonite. The $_{xc}[CO_3^{2-}]$ is an absolute measure of the tendency for the aragonite mineral to precipitate or dissolve[33], so that, unlike the widely used Ω_A (which is a ratio), $_{xc}[CO_3^{2-}]$ can be used to compute $_{xc}[CO_3^{2-}]$ transport. In addition, changes in $_{xc}[CO_3^{2-}]$ are directly comparable with changes in $[CO_3^{2-}]_{is}$, which we computed here from the total alkalinity (A_T) and total dissolved inorganic carbon (C_T). We selected the acid dissociation constants of ref. 34 and aragonite solubilities of ref. 35 to determine $_{xc}[CO_3^{2-}]$ and other derived variables of the marine CO_2 system using the CO2SYS toolbox[36,37].

Present, natural and projected values of $_{xc}[CO_3^{2-}]$ and $[H^+]$. We determined present-day $_{xc}[CO_3^{2-}]$ and $[H^+]$ by using measured data from the marine CO_2 system (A_T, C_T and pH), as well as temperature, salinity and pressure. We separated anthropogenic from natural (preindustrial) $_{xc}[CO_3^{2-}]$ and $[H^+]$ by assuming that the natural component corresponds to an ocean in equilibrium with a preindustrial atmosphere, whereas the remaining is anthropogenic. Along the Ovide section, we determined the anthropogenic CO_2 (C_{ant}) using the back-calculation φC_T° method[19,38,39] with an overall uncertainty of $\pm 5.2\,\mu mol\,kg^{-1}$. For the Nordic Sills, we took the parameters of the marine CO_2 system from ref. 18.

We determined the natural C_T by subtracting the C_{ant} from the measured C_T. We determined natural $_{xc}[CO_3^{2-}]$ and $[H^+]$ using the measured A_T and natural C_T and the equations of the marine CO_2 system using CO2SYS toolbox[36,37].

The shared socioeconomic pathways (specifically SSP; ref. 4; Extended Data Fig. 2) indicate that a doubling of the present atmospheric C_{ant} of 100–120 p.p.m.—that is, an atmospheric CO_2 level of 480–520 p.p.m., which is 200–240 p.p.m. over preindustrial values—would lead to a warming of 2 °C. We calculated $_{xc}[CO_3^{2-}]$ and $[H^+]$ for this scenario by doubling the present atmospheric C_{ant}. Comparing the observed changes in C_T and chlorofluorocarbons with those predicted from an eddy-permitting ocean circulation model confirmed that changes in C_{ant} follow a transient steady-state (TSS) approximation[22,23,40]. This means that C_{ant} increases over time through the whole water column proportionally to the increase in C_{ant} in the surface mixed layer. Hence, we estimated the seawater C_{ant} that would result from a doubling of atmospheric C_{ant} by using the present-day seawater C_{ant}, considering that the change in the seawater surface C_{ant} follows the change in atmospheric CO_2. In this scenario, we estimated the seawater C_{ant} to be $72 \pm 3\%$ higher than the present-day seawater C_{ant} (Extended Data Fig. 4), given a large range of temperatures (0–20 °C) and a CO_2 air–sea disequilibrium[41] of $10 \pm 5\%$. We determined $_{xc}[CO_3^{2-}]$ and $[H^+]$ in the scenario involving a doubling of atmospheric C_{ant} by adding the 0.72 times the present seawater C_{ant} to the measured C_T and using the measured A_T and temperature. $[H^+] = 10^{-pH_T}$ is given in nmol kg^{-1} ('T' refers to pH in total scale). We determined the uncertainty range for the ASH given a doubling of atmospheric C_{ant} (grey area in Fig. 1c) by considering an atmospheric CO_2 concentration of 480–520 p.p.m.

Present, natural and projected $_{xc}[CO_3^{2-}]$ at CWC locations. We derived a database of information on stony coral from literature records[42,43] and from a database on *L. pertusa* that was compiled by the United Nations Environment Programme (UNEP)–World Conservation Monitoring Centre (WCMC)[44]. From the resulting 6,553 records on CWC, we removed those without geo-referenced coordinates or depth data, those referring to depths shallower than 1,000 metres, and those characterized as dead or fossils, thereby obtaining 548 CWC records (Fig. 1a and Extended Data Table 1). To determine the physicochemical characteristics of the seawater surrounding those 548 locations, we used a multiparametric method (the water mass properties (WMP) method)[45] to interpolate *in situ* temperature, salinity, A_T and C_T from the quality-controlled GLODAPv2 database[12,13]. The WMP method improves the estimates obtained from purely spatial interpolation by applying a weighting process that uses the distances in conservative parameters, and by using a helper database with higher resolution (World Ocean Atlas[46–49]) for these conservative parameters. We interpolated the C_{ant} at those 548 CWC locations for 2010 from the C_{ant} given in ref. 50. Following the same procedure as described above, we determined present-day, natural and projected $_{xc}[CO_3^{2-}]$ (the latter for a doubling of atmospheric C_{ant}). The results are summarized in Extended Data Table 1.

Circulation of CO_3^{2-} and C_{ant} in the subpolar North Atlantic. We determined the absolute geostrophic velocity fields orthogonal to the Ovide sections by using an inverse model constrained by subsurface acoustic Doppler current profiler measurements and an overall mass conservation constraint[17,51–53]. The resulting absolute velocity fields are consistent with independent altimetry measurements[52] and estimates of the western boundary current transport[53] at the time of the Greenland–Portugal oceanographic cruises (made every two years since 2002) along the Ovide transect (Fig. 1b). The velocity fields are representative of the month of the cruise; seasonal variability was removed as in ref. 19. The mean velocity field of the Ovide section during 2002–2016 is shown in Extended Data Fig. 5. We calculated the transport of $_{xc}[CO_3^{2-}]$, $[H^+]$, C_{ant} and $[CO_3^{2-}]$ by using the velocity field, density and tracer concentration ($_{xc}[CO_3^{2-}]$, $[H^+]$, C_{ant} or $[CO_3^{2-}]$) as follows:

$$\text{Transport}_{\text{tracer}} = \Sigma\,[\text{tracer}] \times \text{density}_{\text{is}} \times \text{velocity} \times \text{area}$$

The transports of seawater, $_{xc}[CO_3^{2-}]$, $[H^+]$, C_{ant} and $[CO_3^{2-}]$ are summarized in Extended Data Table 2, Fig. 3 and Extended Data Fig. 6. For $_{xc}[CO_3^{2-}]$ and $[H^+]$, transports are also given for the present day, for preindustrial (natural) times, and for a scenario involving doubling of atmospheric C_{ant}. We calculated uncertainties in the AMOC, and in the transport of $_{xc}[CO_3^{2-}]$ and $[H^+]$ through the Ovide section, as the standard errors of the mean of the eight occupations of the Ovide section during 2002–2016. Transport at the sills was derived from refs 18, 19; results are summarized in Extended Data Table 3.

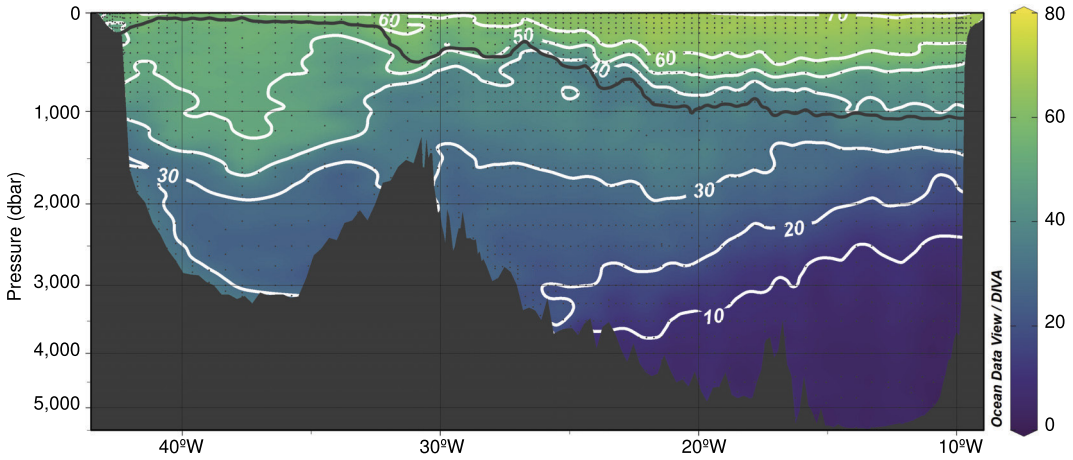
Data availability. Data obtained by the OVIDE cruises from 2002 to 2010 are part of the GLODAPv2 data product, available at Ocean Carbon Data System (http://cdiac.ornl.gov/oceans). We accessed OVIDE 2012 data from the Clivar and Carbon Hydrographic Data Office (CCHDO; http://cchdo.ucsd.edu/cruise/29AH20120622). OVIDE 2014 data are publicly available at http://www.obs-vlfr.fr/proof/ftpfree/geovide/ALKALINITY_PH/. The OVIDE 2016 data that support the findings of this study are available at http://dx.doi.org/10.20350/digitalCSIC/8513.

30. Fabry, V. J., Seibel, B. A., Feely, R. A. & Orr, J. C. Impacts of ocean acidification on marine fauna and ecosystem processes. *ICES J. Mar. Sci.* **65,** 414–432 (2008).
31. Fabry, V. O., McClintock, J. B., Mathis, J. T. & Grebmeier, J. M. Ocean acidification at high latitudes: the bellwether. *Oceanography* **22,** 160–171 (2009).
32. Millero, F. J., Feistel, R., Wright, D. G. & McDougall, T. J. The composition of standard seawater and the definition of the reference-composition salinity scale. *Deep Sea Res. I* **55,** 50–72 (2008).
33. Subhas, A. V. *et al.* A novel determination of calcite dissolution kinetics in seawater. *Geochim. Cosmochim. Acta* **170,** 51–68 (2015).
34. Dickson, A. & Millero, F. A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media. *Deep Sea Res. A* **34,** 1733–1743 (1987).
35. Mucci, A. The solubility of calcite and aragonite in seawater at various salinities, temperatures, and one atmosphere total pressure. *Am. J. Sci.* **283,** 780–799 (1983).
36. Lewis, E., Wallace, D. & Allison, L. *Program Developed For CO_2 System Calculations* (Carbon Dioxide Information Analysis Center, U.S. Department of Energy, 1998).
37. van Heuven, S., Pierrot, D., Lewis, E. & Wallace, D. *MATLAB Program Developed for CO_2 System Calculations* (Carbon Dioxide Information Analysis Center, U.S. Department of Energy, 2009).
38. Pérez, F. F. *et al.* Temporal variability of the anthropogenic CO_2 storage in the Irminger Sea. *Biogeosciences* **5,** 1669–1679 (2008).
39. Vázquez-Rodríguez, M. *et al.* Anthropogenic carbon distributions in the Atlantic Ocean: data-based estimates from the Arctic to the Antarctic. *Biogeosciences* **6,** 439–451 (2009).
40. Tanhua, T., Körtzinger, A., Friis, K., Waugh, D. W. & Wallace, D. W. R. An estimate of anthropogenic CO_2 inventory from decadal changes in oceanic carbon content. *Proc. Natl Acad. Sci. USA* **104,** 3037 (2007).
41. Matsumoto, K. & Gruber, N. How accurate is the estimation of anthropogenic carbon in the ocean? An evaluation of the ΔC* method. *Glob. Biogeochem. Cycles* **19,** GB3014 (2005).
42. Rogers, A. D., Baco, A., Griffiths, H., Hart, T. & Hall-Spencer, J. M. in *Seamounts: Ecology, Fisheries and Conservation* (eds Pitcher, T. J. *et al.*) 141–169 (Blackwell, 2007).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

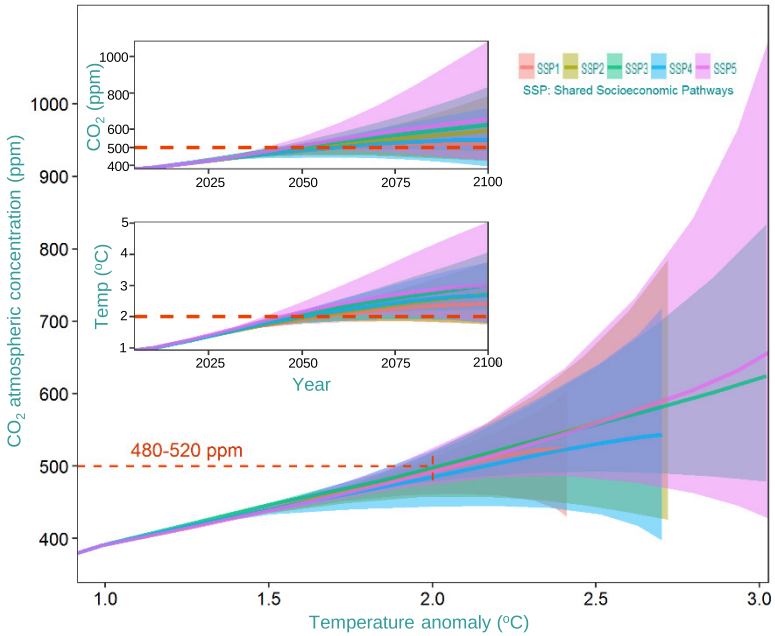
43. Hall-Spencer, J., Rogers, A., Davies, J. & Foggo, A. Deep-sea coral distribution on seamounts, oceanic islands, and continental slopes in the Northeast Atlantic. *Bull. Mar. Sci.* **81,** 135–146 (2007).
44. Freiwald, A. *et al.* Global distribution of cold-water corals (version 3.0). *UNEP World Conservation Monitoring Centre* http://data.unep-wcmc.org/datasets/3. (2017).
45. Velo, A. *et al.* A multiparametric method of interpolation using WOA05 applied to anthropogenic CO_2 in the Atlantic. *Sci. Mar.* **74,** 21–32 (2010).
46. Locarnini, R. A. *et al. World Ocean Atlas 2013. Volume 1: Temperature* (NOAA Atlas NESDIS, 2013).
47. Zweng, M. *et al. World Ocean Atlas 2013. Volume 2: Salinity* (NOAA Atlas NESDIS, 2013).
48. Garcia, H. E. *et al. World Ocean Atlas 2013. Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation.* (NOAA Atlas NESDIS, 2014).
49. Garcia, H. E. *et al. World Ocean Atlas 2013, Volume 4: Dissolved Inorganic Nutrients (Phosphate, Nitrate, Silicate).* (NOAA Atlas NESDIS, 2014).
50. Khatiwala, S., Primeau, F. & Hall, T. Reconstruction of the history of anthropogenic CO_2 concentrations in the ocean. *Nature* **462,** 346–349 (2009).
51. Lherminier, P. *et al.* The Atlantic meridional overturning circulation and the subpolar gyre observed at the A25-OVIDE section in June 2002 and 2004. *Deep Sea Res. I* **57,** 1374–1391 (2010).
52. Gourcuff, C., Lherminier, P., Mercier, H. & Le Traon, P. Y. Altimetry combined with hydrography for ocean transport estimation. *J. Atmos. Ocean. Technol.* **28,** 1324–1337 (2011).
53. Daniault, N., Lherminier, P. & Mercier, H. Circulation and transport at the southeast tip of Greenland. *J. Phys. Oceanogr.* **41,** 437–457 (2011).
54. Kriegler, E. *et al.* Fossil-fueled development (SSP5): an energy and resource intensive scenario for the 21st century. *Glob. Environ. Change* **42,** 297–315 (2017).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



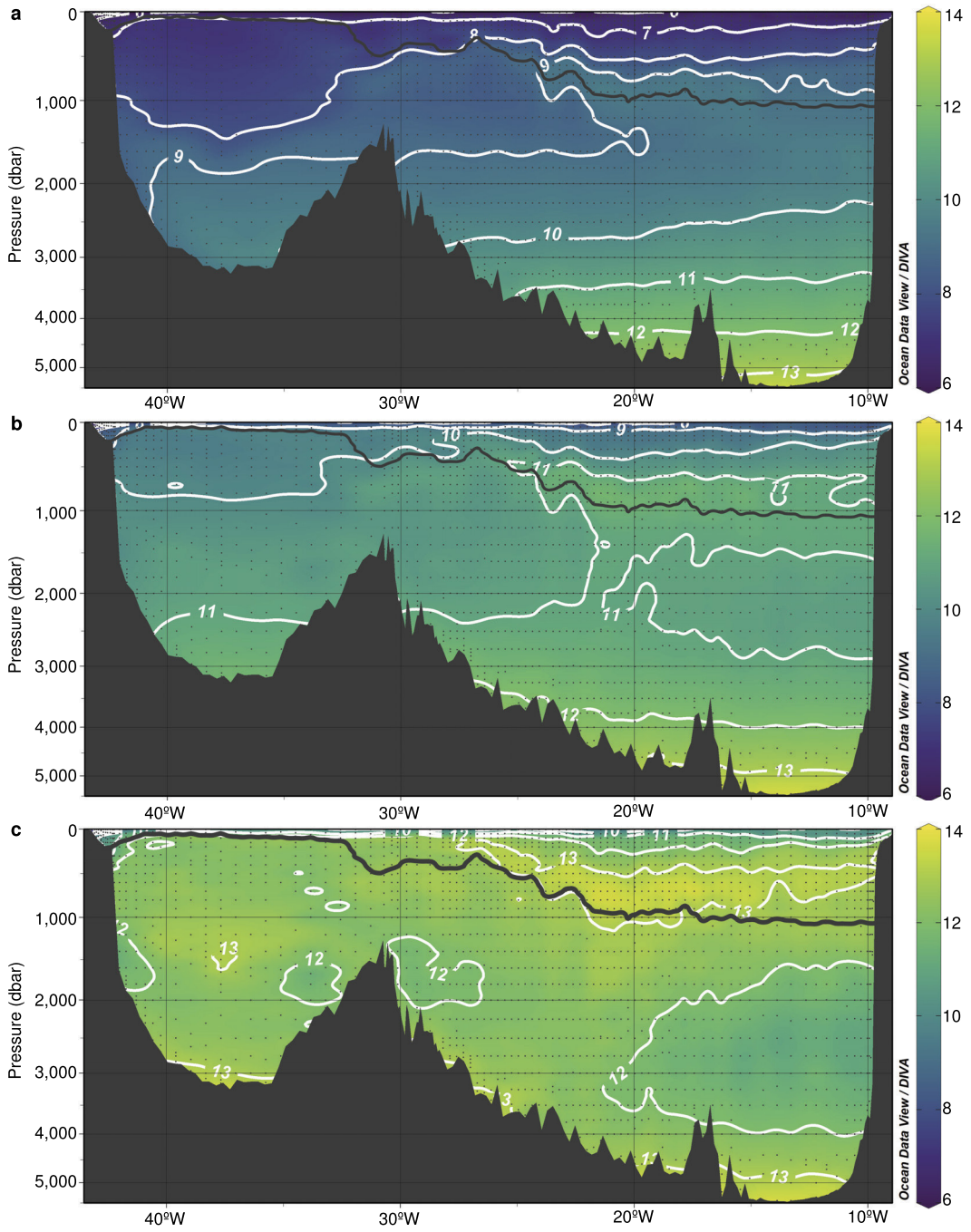
Extended Data Figure 1 | Anthropogenic CO_2 along the 2016 Ovide section. Vertical distribution of anthropogenic CO_2 (colour bar and white isolines; in μmol kg^{-1}) along the Ovide section of the North Atlantic (Fig. 1b) in 2016. The black line marks the isopycnal that delimits the upper and lower limbs of the AMOC. Dots represent samples. Dark shaded area represents the sea bed.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



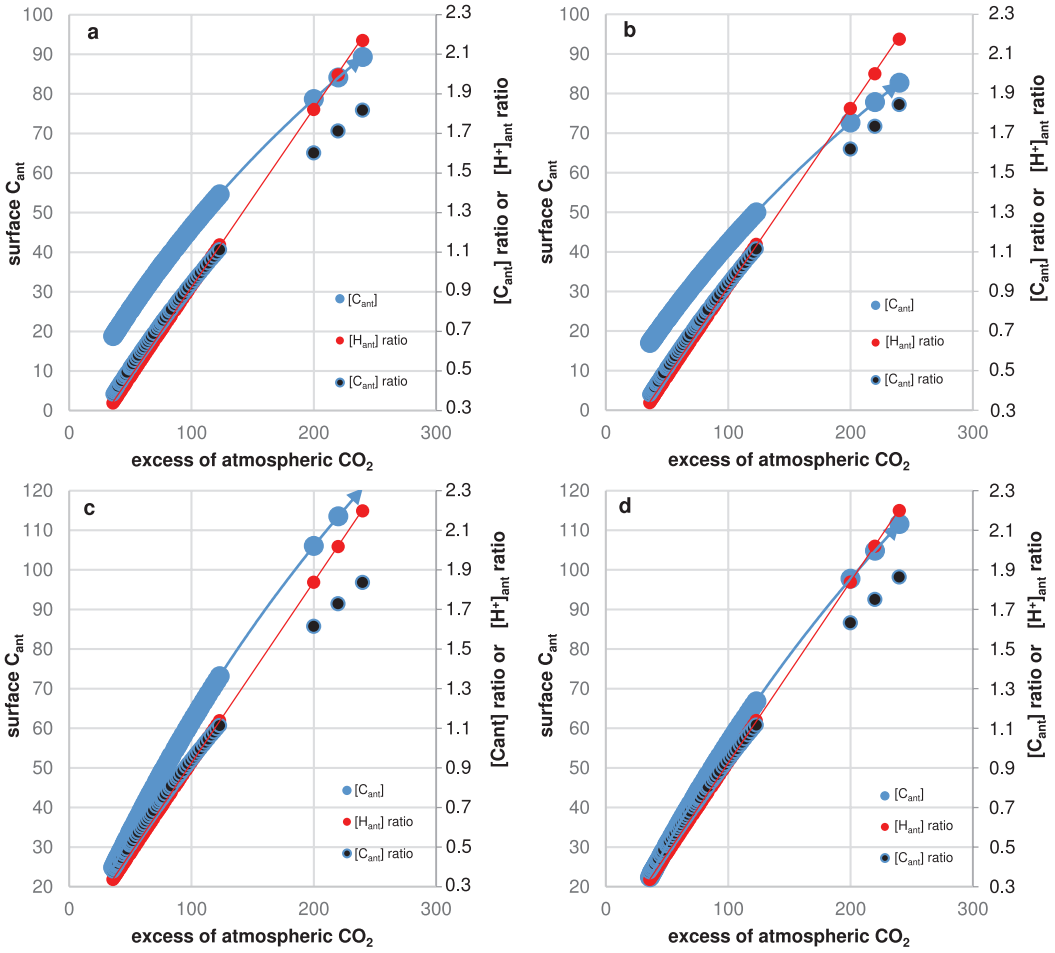
Extended Data Figure 2 | Projections of atmospheric CO_2 and temperature changes as modelled by the five shared socioeconomic pathways (SSPs)[24]. At present (bottom left of each graph), Earth has experienced a 1 °C warming with respect to preindustrial temperatures, and has an atmospheric CO_2 concentration ($[CO_2^{atm}]$) of 400 p.p.m. (ref. 21); this represents an atmospheric CO_2 excess ($_{xc}[CO_2^{atm}]$) of 120 p.p.m. with respect to preindustrial times ($_{xc}[CO_2^{atm}] = [CO_2^{atm}] - 280$ p.p.m., where 280 p.p.m. is the preindustrial $[CO_2^{atm}]$). If atmospheric C_{ant} were to double, so too would $_{xc}[CO_2^{atm}]$ (to 480–520 p.p.m.; dotted red line in main graph); this is within the projection range for a warming of 2 °C. According to SSP5 (pink shaded area), this temperature could be reached in 2040–2052 (see insets). SSP5 is the only one of these five scenarios that results in a radiative forcing as high as that projected by the previous IPCC business-as-usual scenario (RCP8.5)[54].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



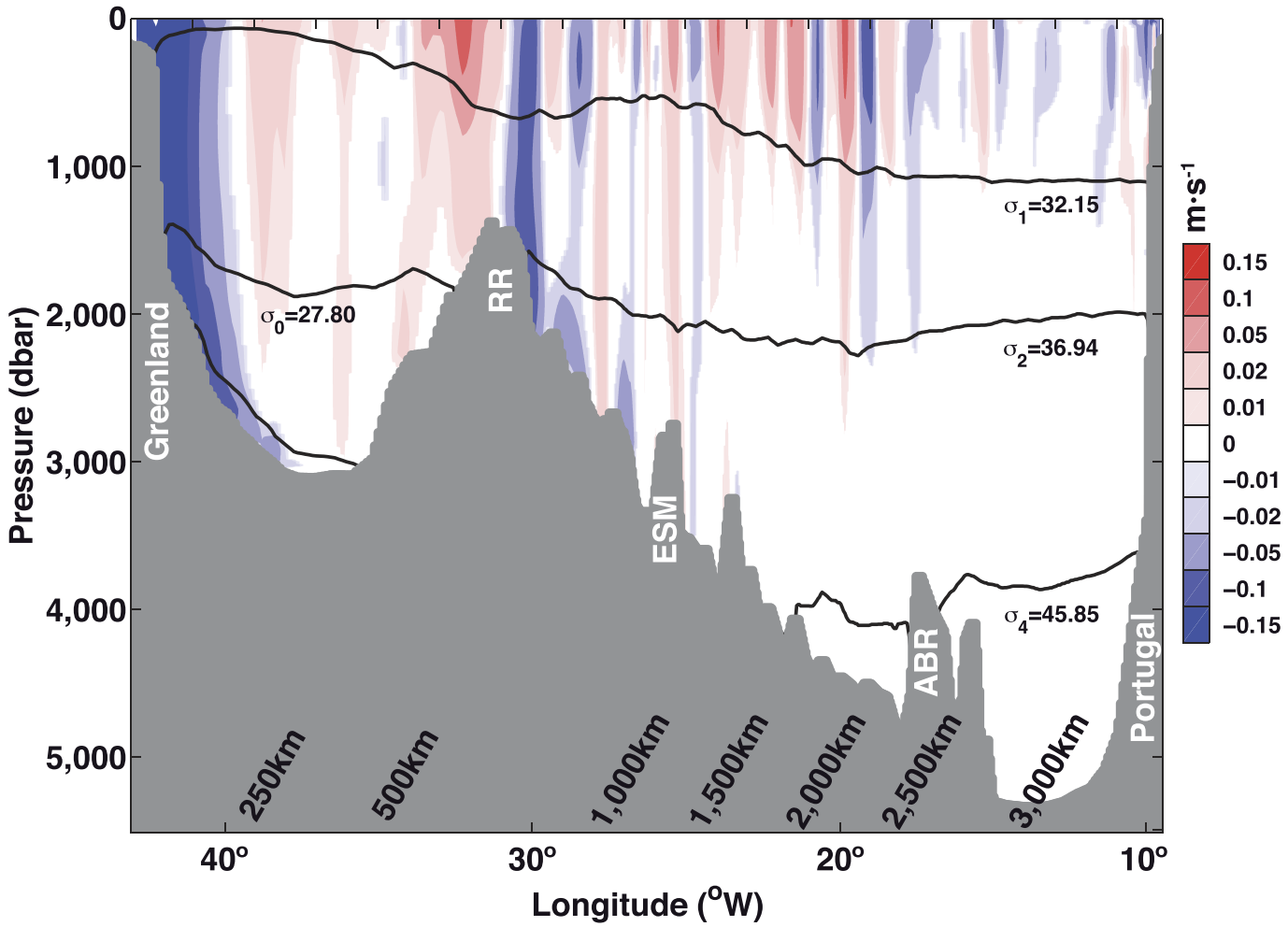
Extended Data Figure 3 | Evolution of ocean acidification from preindustrial to the doubling of atmospheric C_{ant} scenario. a–c, Along the Ovide section (Fig. 1b) are shown H^+ concentrations (colour bar and white isolines; in nmol kg^{-1}) for preindustrial times (colour bar and white isolines) (**a**), the present day (**b**; 2016), and in a scenario involving doubling of atmospheric C_{ant} (**c**). The black line shows the isopycnal delimiting the upper and lower limbs of the AMOC. Dots represent samples. Dark shaded area denotes the sea bed.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



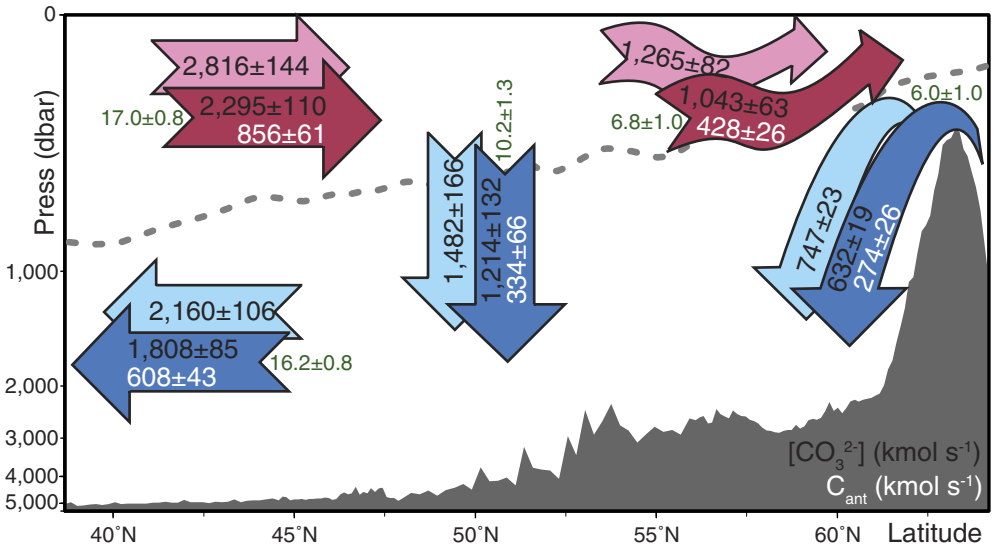
Extended Data Figure 4 | Progression of anthropogenic CO_2 in surface waters versus excess of atmospheric CO_2 at different temperatures and CO_2 disequilibriums. a–d, The concentration of anthropogenic CO_2 ($[C_{ant}]$, in μmol kg^{-1}; blue dots), the C_{ant} ratio (navy encircled dots) and the H_{ant} ratio (red dots) are plotted against the excess of atmospheric CO_2 that has resulted from anthropogenic activity during 1959–2016 for: **a**, a 2 °C warming and 5% of CO_2 disequilibrium; **b**, a 2 °C warming and 15% of CO_2 disequilibrium; **c**, a 20 °C warming and 5% of CO_2 disequilibrium; and **d**, a 20 °C warming and 15% of CO_2 disequilibrium. The C_{ant} ratio (or H_{ant} ratio) is the ratio between the $[C_{ant}]$ (or $[H^+]_{ant}$) at any time and the average $[C_{ant}]$ ($[H^+]_{ant}$) for the period covered by the OVIDE cruises (2002–2016). The values computed for the doubling of atmospheric C_{ant} scenario (that is, for an excess of atmospheric CO_2 of 220 ± 20 p.p.m.) are also shown (blue arrowheads). The progression of the H_{ant} ratio fits to a straight line, while the C_{ant} ratio decreases when increasing the atmospheric excess CO_2 owing to the buffer capacity of the marine CO_2 system (the Revelle factor). An increase of 72 ± 3% in C_{ant} above the average $[C_{ant}]$ for the Ovide period (2002–2016) is computed for the doubling of atmospheric C_{ant} scenario. This is practically independent of the seawater temperature and the degree of CO_2 disequilibrium[41].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 5 | Mean velocities orthogonal to Ovide. Mean velocities (in m s^{-1}) orthogonal to the Ovide section for the period 2002–2016. The Ekman velocities—estimated from wind stress averaged over the months of the cruises and equally distributed over the first 30 m—were added to the geostrophic velocities derived from a box inverse model[17,51–53]. Positive (or negative) velocities indicate northeastward (or southwestward) currents. The horizontal grey lines are the isopycnals marking $\sigma_0 = 27.80\,kg\,m^{-3}$, $\sigma_1 = 32.15\,kg\,m^{-3}$, $\sigma_2 = 36.94\,kg\,m^{-3}$ and $\sigma_4 = 45.85\,kg\,m^{-3}$. At the bottom are shown distances from Greenland. Grey shaded area denotes the sea bed. The acronyms indicate topographic features: the Reykjanes Ridge (RR), the Eriador Seamount (ESM), and the Azores–Biscay Ridge (ABR).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 6 | Circulation of CO_3^{2-} ions and anthropogenic CO_2 in the subpolar North Atlantic. Mean transports for 2002–2016 between the Ovide section and the Nordic Sills are shown for: sea water (green values, in Sv), as well as natural (preindustrial; light coloured arrows) and observed (dark coloured arrows) [CO_3^{2-}] (black values) and C_{ant} (white values) (both in kmol s^{-1}). Uncertainties are the errors of the mean transports across the eight occupations of the Ovide section. Transports at the Nordic Sills are derived from refs 18, 19. The grey dotted line represents the limit between the upper and lower limbs of the AMOC. Grey shaded area denotes the sea bed.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Aragonite saturation at CWC locations

Depth (m)	Ocean	N	$<ASH_{nat}$	$<ASH_{2010}$	$<ASH_{2x}$	$xc[CO_3^{2-}]_{nat}$	$xc[CO_3^{2-}]_{2010}$	$xc[CO_3^{2-}]_{2x}$
≥1,000	Global	548	12%	17%	47%	29	14.5	4
	NA	335(61%)	1%	2%	32%	43	24.5	10.5
≥1,500	Global	176	20%	22%	74%	22	6.7	-8
	NA	139(78%)	2%	4%	70%	34	15	-3

Number (N) of locations of CWC in the global and North Atlantic (NA) oceans at depths below 1,000 metres and 1,500 metres, along with the percentage of the CWCs located below the ASH (<ASH) in preindustrial (natural) times (subscript 'nat'), in the present day (subscript '2010'), and in a scenario involving the doubling of atmospheric C_{ant} (subscript '2x'). Also shown are average values for the excess of CO_3^{2-} concentration over aragonite saturation ($_{xc}[CO_3^{2-}]$, in μmol kg^{-1}) for preindustrial times, for the present day (from the GLODAPv2 project)[12,13], and with a doubling of atmospheric C_{ant}.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Transport of sea water, $_{xc}[CO_3^{2-}]$, $[H^+]$, C_{ant} and $[CO_3^{2-}]$ across the Ovide section

Year	Vol (Sv)	C_{ant} (kmol·s^{-1})	$[CO_3^{2-}]$ (kmol·s^{-1})	$[CO_3^{2-}]_{nat}$ (kmol·s^{-1})	$_{xc}[CO_3^{2-}]$ (kmol·s^{-1})	$_{xc}[CO_3^{2-}]_{nat}$ (kmol·s^{-1})	$_{xc}[CO_3^{2-}]_{2x}$ (kmol·s^{-1})	$[H^+]$ (pmol·s^{-1})	$[H^+]_{nat}$ (pmol·s^{-1})	$[H^+]_{2x}$ (pmol·s^{-1})
AMOC upper limb										
2002	17.3	761	2,322	2,784	1,058	1,520	737	177	144	209
2004	17.2	820	2,409	2,910	1,172	1,673	823	172	138	205
2006	12.0	528	1,596	1,917	743	1,064	520	119	97	141
2008	17.1	886	2,412	2,954	1,173	1,716	795	171	135	206
2010	18.0	919	2,477	3,037	1,166	1,727	776	187	148	225
2012	16.5	819	2,116	2,609	922	1,415	578	172	135	208
2014	18.8	1,048	2,563	3,200	1,199	1,837	755	189	147	230
2016	19.0	1,068	2,468	3,114	1,078	1,724	628	201	154	247
Avg.	**17.0**	**856**	**2,295**	**2,816**	**1,064**	**1,584**	**701**	**174**	**137**	**209**
Err_x	**0.78**	**61**	**110.3**	**144.2**	**56**	**88**	**39**	**9**	**6**	**11**
Velocity-weighted averaged $_{xc}[CO_3^{2-}]$					**61±2**	**90±2**	**40±1.4**			
AMOC lower limb										
2002	-16.5	-548	-1,920	-2,240	-640	-960	-417	-151	-127	-175
2004	-16.3	-547	-1,812	-2,128	-454	-770	-233	-156	-130	-181
2006	-11.2	-343	-1,294	-1,495	-357	-557	-218	-106	-90	-121
2008	-16.3	-694	-1,899	-2,305	-680	-1,087	-397	-150	-120	-180
2010	-17.2	-567	-1,913	-2,239	-288	-614	-60	-175	-147	-202
2012	-15.7	-646	-1,769	-2,144	-524	-899	-263	-151	-121	-181
2014	-17.9	-636	-1,899	-2,262	-259	-622	-6	-184	-151	-215
2016	-18.2	-882	-1,957	-2,465	-500	-1,007	-146	-184	-140	-226
Avg.	**-16.2**	**-608**	**-1,808**	**-2,160**	**-463**	**-814**	**-217**	**-157**	**-128**	**-185**
Err_x	**0.78**	**43**	**84.8**	**106**	**55**	**73**	**46**	**8**	**7**	**10**
Velocity-weighted averaged $_{xc}[CO_3^{2-}]$					**28±3**	**49±4**	**13.6±2.7**			

Shown are measured values for the present-day transport of seawater volume, C_{ant}, $[CO_3^{2-}]$, $_{xc}[CO_3^{2-}]$ and $[H^+]$ across the Ovide section during 2002–2016 in the upper and lower limbs of the AMOC. Also shown are values for the natural (preindustrial) components (subscript 'nat') and for the scenario involving doubling of atmospheric C_{ant} (subscript '2x'). Uncertainties (*Err_x*) are estimated as the error of the mean of the eight occupations of the Ovide section. Positive (negative) transports indicate northward (southward) transport. Also shown are velocity-weighted averages of $_{xc}[CO_3^{2-}]$ for preindustrial times, the present day (2002–2016), and for a doubling of atmospheric C_{ant} scenario.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 3 | Transports at the Nordic Sills

Layer	Vol (Sv)	T_{pot} (°C)	Salinity	C_T (μmol·kg^{-1})	A_T (μmol·kg^{-1})	C_{ant} (μmol·kg^{-1})	$[CO_3^{2-}]$ (μmol·kg^{-1})	$[CO_3^{2-}]_{nat}$ (μmol·kg^{-1})	$_{xc}[CO_3^{2-}]$ (μmol·kg^{-1})	$_{xc}[CO_3^{2-}]_{nat}$ (μmol·kg^{-1})	$[H^+]$ (pmol·kg^{-1})	$[H^+]_{nat}$ (pmol·kg^{-1})
AMOC upper limb (Density < 1,027.8)												
$AW_{Faroes\text{-}Shetland}$	3.9±0.5	9.18±0.55	35.27±0.11	2,121±20	2,325±7	50±3	146±8	176±10	76±8	107±10	8.0±0.9	6.4±1.0
$AW_{Iceland\text{-}Faroe}$	3.9±0.5	8.10±0.30	35.18±0.06	2,127±22	2,323±6	47±3	140±10	169±12	70±10	99±12	8.0±1.1	6.4±1.3
$AW_{W\text{-}Iceland}$	0.8±0.16	6.00±0.15	35.06±0.03	2,138±3	2,309±2	48±2	124±2	153±2	54±2	83±2	8.4±0.1	6.6±0.2
$PW_{Denmark\ Strait}$	-1.8±0.5	-0.91±1.35	34.07±0.27	2,114±8	2,266±3	38±1	110±3	132±4	42±3	64±4	6.9±0.4	5.6±0.5
AMOC lower limb (Density > 1,027.8)												
$OW_{Faroes\text{-}Shetland}$	-2.0±0.3	0.59±0.10	34.91±0.02	2,163±3	2,303±3	27±6	100±2	115±4	2±2	17±4	8.1±0.0	7.0±0.5
$OW_{Iceland\text{-}Faroe}$	-1.0±0.5	0.33±0.30	34.89±0.06	2,156±5	2,304±8	33±4	104±2	123±5	6±2	25±5	7.7±0.2	6.4±0.1
$OW_{Denmark\ Strait}$	-3.0±0.3	1.47±0.45	34.86±0.09	2,148±6	2,294±5	37±2	104±2	124±3	6±2	27±3	8.1±0.1	6.6±0.2

Shown are seawater volume transport (Vol), potential temperature (Tpot), salinity, total dissolved inorganic carbon (C_T), total alkalinity (A_T), C_{ant}, $[CO_3^{2-}]$, $_{xc}[CO_3^{2-}]$ and $[H^+]$ through the Nordic Sills in the upper and lower limbs of the AMOC (separated by isopycnal 1,027.8 kg m^{-3}). Values for certain natural (preindustrial) components are also shown (subscript 'nat'). Salinity and marine CO_2 system data (C_T and A_T) are from ref. 17. Volume transports are from ref. 19. AW, OW and PW: Atlantic, overflow and polar waters. Faroes–Shetland, between the Faroes and Shetland Islands; Iceland–Faroe, across the Iceland–Faroe Ridge; W-Iceland, west of Iceland; Denmark Strait, through the Denmark Strait. Uncertainties in volume transport, salinity, C_T, C_{ant} and A_T are from ref. 18. Uncertainties in C_T and A_T were propagated to $[CO_3^{2-}]$ and $[H^+]$ by a Monte Carlo approach using marine CO_2 system equations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature25508

Global patterns of tropical forest fragmentation

Franziska Taubert[1], Rico Fischer[1], Jürgen Groeneveld[1,2], Sebastian Lehmann[1], Michael S. Müller[1], Edna Rödig[1], Thorsten Wiegand[1,3] & Andreas Huth[1,3,4]

Remote sensing enables the quantification of tropical deforestation with high spatial resolution[1,2]. This in-depth mapping has led to substantial advances in the analysis of continent-wide fragmentation of tropical forests[1–4]. Here we identified approximately 130 million forest fragments in three continents that show surprisingly similar power-law size and perimeter distributions as well as fractal dimensions. Power-law distributions[5–7] have been observed in many natural phenomena[8,9] such as wildfires, landslides and earthquakes. The principles of percolation theory[7,10,11] provide one explanation for the observed patterns, and suggest that forest fragmentation is close to the critical point of percolation; simulation modelling also supports this hypothesis. The observed patterns emerge not only from random deforestation, which can be described by percolation theory[10,11], but also from a wide range of deforestation and forest-recovery regimes. Our models predict that additional forest loss will result in a large increase in the total number of forest fragments—at maximum by a factor of 33 over 50 years—as well as a decrease in their size, and that these consequences could be partly mitigated by reforestation and forest protection.

Tropical forests have a key role in the global carbon cycle[12] and harbour more than half of the known species worldwide[13]. In recent decades, increases in agriculture, logging and urban growth have led to unprecedented losses of tropical forest[14,15], with annual deforestation rates of around 0.5% since the 1990s[1]. Deforestation rates differ between continents[1], with hot spots concentrated mainly in Asia and Brazil[14]. A reduction of forested area is concomitant with fragmentation, in which patches of forest are split into several smaller ones[16]. The increasing availability of high-resolution satellite imagery now enables in-depth mapping of global deforestation[1–4] and detailed analysis of the spatial pattern of the remaining forest.

Here we used percolation theory[10,11] as a framework for analysing the current fragmentation structures in tropical and subtropical regions of the Americas, Africa, and Asia and Australia (denoted here as Asia–Australia) (Fig. 1) on a high-resolution forest cover map[2] (approximately 21 billion pixels, each 30 m × 30 m). We used a clustering algorithm to count and analyse the size and perimeter distribution of all detectable forest fragments separately for each continent, and determined the fractal dimension of the fragmented landscape (Table 1). In total, we identified more than 130 million forest fragments across all continents studied, with sizes ranging over eleven orders of magnitude and reaching 427 Mha. South America's largest forest fragment, in the Amazon, spans around 45% of its total forest area, whereas the largest fragment on Borneo in Asia covers only 18% of the forest.

We expected notably different fragmentation structures between continents owing to different land use practices[14]. However, we observed markedly similar fragment size distributions (Fig. 1), which could be described by power laws[5–7] with almost identical exponents (Table 1). About 10% of continental forest area is made up of forest fragments smaller than 10,000 ha (11.2% for the Americas, 9.9% for Africa and 9.2% for Asia–Australia). Additionally, the fractal dimensions of forest cover are similar for all three continents, with values of approximately 1.9 (Table 1). To analyse the shape of the fragments, we determined the fragment perimeter distributions; again, for all three continents, we found power-law behaviour with similar exponents, with values close to 2.2 (Table 1, Extended Data Fig. 1). Our results raise the

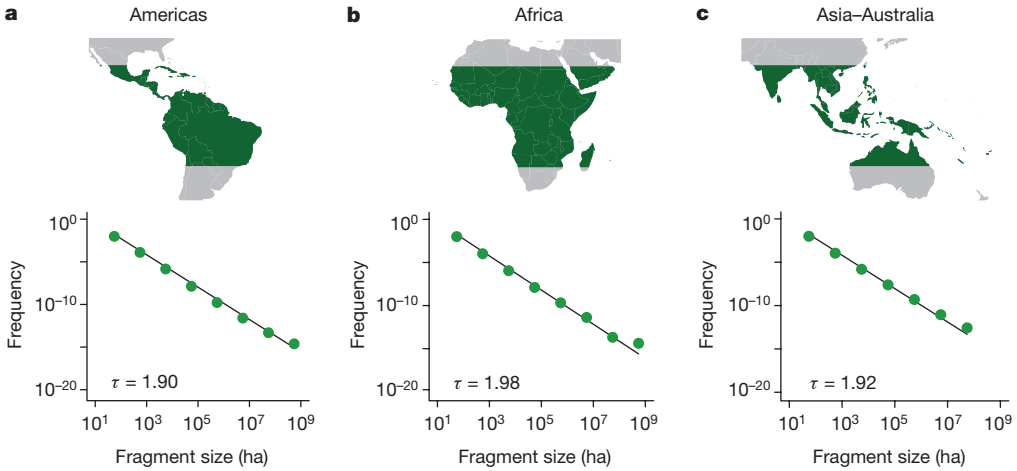


Figure 1 | Continental-scale fragment size distribution of tropical and subtropical forests. a–c, Observed forest fragment size distribution (green dots, fragment sizes ≥10 ha) for the Americas (**a**, $n = 55.5$ million fragments), Africa (**b**, $n = 44.8$ million fragments) and Asia–Australia (**c**, $n = 30.5$ million fragments). The solid lines represent the fits of power-law distributions with exponent τ. The maps show the entire land area of the tropical belt and indicate the selected tropical regions in green. Maps were created using R (ref. 31) and the package 'rworldmap' (ref. 32). Fragment sizes were estimated from Hansen's vegetation cover map[2].

[1]Helmholtz Centre for Environmental Research – UFZ, Department of Ecological Modelling, Permoserstrasse 15, 04318 Leipzig, Germany. [2]TU Dresden, Institute of Forest Growth and Forest Computer Sciences, PO 1117, 01735 Tharandt, Germany. [3]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany. [4]University of Osnabrück, Institute of Environmental Systems Research, Barbarastrasse 12, 49076 Osnabrück, Germany.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Table 1 | Continental-scale fragmentation of tropical forests

	Americas	Africa	Asia–Australia	Theory*
Forest area in Mha	940	577	391	
Number of fragments (n)	55,558,018	44,851,251	30,556,204	
Mean (median) size of fragments in ha	17 (0.09)	13 (0.09)	13 (0.09)	
Forest area (%) of fragments <10,000 ha	11.2	9.9	9.2	
Fragment size scaling†,‡ (τ)	1.90	1.98	1.92	2.05
Fractal dimension§ (d_f)	1.92	1.87	1.87	1.89
Perimeter scaling†,‖ (κ)	2.16	2.23	2.21	2.14

All features were calculated on the basis of a high-resolution forest cover map[2]. Annual gross deforestation rates[1] of 0.51% (Americas), 0.37% (Africa) and 0.62% (Asia–Australia) for the years 2000–2010.
*Theoretical values according to percolation theory[10,11] (near the critical point).
†The term 'scaling' refers to the exponent of the fitted power-law distributions. The calculation of fragment size and perimeter scaling is based on $n = 55.5$ million (Americas), $n = 44.8$ million (Africa) and $n = 30.5$ million (Asia–Australia) fragments. See Methods for details.
‡Pearson's correlation coefficients (R) and standard errors (ε) are: $R^2 = 0.96$ and $\varepsilon = 0.0011$ (Americas), $R^2 = 0.956$ and $\varepsilon = 0.0017$ (Africa), $R^2 = 0.958$ and $\varepsilon = 0.0022$ (Asia–Australia). Deviations from the theoretical expectation trace back to the finite area of the continents.
§The fractal dimension is based on the analysis of $n = 8$ different grid sizes, see Methods for details. The estimate of the fractal dimension is not affected by landscape area.
‖Pearson's correlation coefficients (R) and standard errors (ε) are: $R^2 = 0.952$ and $\varepsilon = 0.00082$ (Americas), $R^2 = 0.931$ and $\varepsilon = 0.0011$ (Africa), $R^2 = 0.938$ and $\varepsilon = 0.0015$ (Asia–Australia).

question of why varying patterns of local deforestation produce fragmentation patterns that are so similar at the continental scale.

Percolation theory[10,11] provides one possible explanation for the observed patterns. In this theory, the probability that the cells of a landscape are occupied by forest is designated p. Where occupied cells share at least one side with another, a forest fragment (also known as a cluster) is formed. If p is large, the landscape is dominated by one large fragment spanning the whole area (Extended Data Fig. 2a). For lower values of p the forest landscape is divided into smaller fragments, although there is still one large fragment (known as the spanning cluster) that connects one side of the landscape with the other. It should be noted that the spanning cluster can form complex shapes with larger holes[11] (Extended Data Fig. 2b). When the probability p falls to a certain level (approximately that of the critical point, $p_c = 0.59$, which indicates 59% forest area for large landscapes), the spanning cluster breaks up and smaller fragments appear (Extended Data Fig. 2c, d).

The phases of fragmentation below and above a forest cover of 59% are defined as subcritical and supercritical, respectively, whereas the value of $p_c = 0.59$ itself represents the critical point of percolation at which the large-scale behaviour of the system can be described by simple mathematical relationships[11,17]. For example, percolation theory predicts that the size distribution of fragments can be described by a power law with an exponent τ of approximately 2.05 at the critical point (for infinite landscapes[10,11]). Notably, the empirical exponents of the forest cover map agree with the theory, with τ values of between 1.9 and 2.0 (Table 1). At the critical point, the exponent of the fragment size distribution τ is further related to the fractal dimension d_f via the hyperscaling relationship[11,18,19] $\tau = 1 + 2/d_f$. Theory predicts a fractal dimension of 1.89 at the critical point[11], which is in agreement with the d_f values determined for the high-resolution forest cover map (1.92 for the Americas and 1.87 for Africa and Asia–Australia, Table 1). The obtained perimeter distributions (Table 1) also show power laws in agreement with the theoretical exponent[20] κ of approximately 2.14. Such critical exponents are often fingerprints of the hidden dynamics of a system[7,9,21].

Our empirical findings suggest that the observed tropical forest fragmentation is near to the critical point of percolation in all three continents (reference year 2000). To test this hypothesis and to explore future dynamics, we developed and analysed three dynamic fragmentation models: FRAG, FRAG-B and FRAG-P. In the FRAG model, which corresponds to classical percolation theory[11], local forest sites are cleared randomly and independently each year[22,23] (Fig. 2, Extended Data Fig. 3, Supplementary Video 1). Deforestation occurs across a landscape initially covered by forest. The FRAG-B and FRAG-P models were used to test our findings with respect to more complex deforestation and reforestation patterns as well as forest protection (see Methods for details). We analysed the dynamics of forest fragmentation for each continent and monitored all forest fragments emerging over time. In total, approximately 3 Gha of forest area was simulated (33 billion cells of resolution 30 m × 30 m).

The dynamics of forest fragmentation as simulated by the FRAG model predict the empirically observed fragment size distributions (Fig. 2b). For forest removal beyond the critical point, the development of the number of fragments shows a 'hump-shaped' behaviour[24] (Fig. 3a, b). The number of fragments increases slowly before the critical point, but accelerates sharply afterwards (to a maximum of 2,000 million fragments for the tropical Americas, Fig. 3a). The number of fragments currently observed and their mean size also match the predictions of the FRAG model close to the critical point (Fig. 3a, b). Small additional amounts of forest loss in the near future would lead to a strong

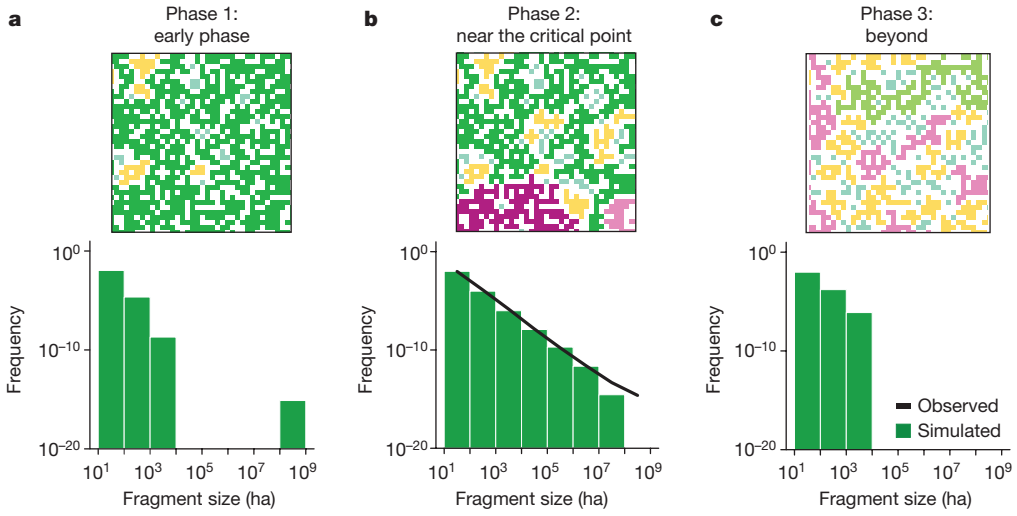


Figure 2 | Dynamics of tropical forest fragmentation in the Americas. **a**–**c**, Fragment size distributions (green bars, FRAG simulation; black line, observations from remote sensing; fragment sizes ≥10 ha) and spatial patterns of fragments at different time points: the early phase (**a**), near to the critical point of percolation (**b**) and beyond (**c**). For each phase, a map of a selected subarea of 225 ha is shown (inset, according to the FRAG model); cleared sites are white and the different colours indicate different fragment sizes. Snapshots and a video are provided in Extended Data Fig. 3 and Supplementary Video 1, respectively.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

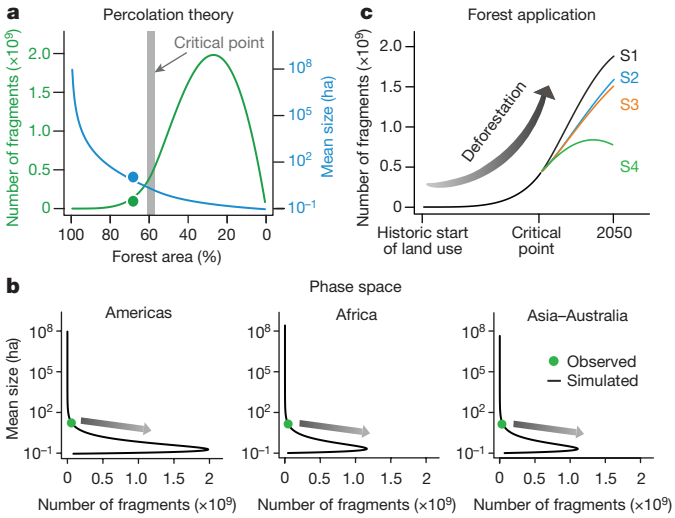


Figure 3 | Percolation theory applied to the observed forest fragmentation. a, Dynamics of the number of forest fragments (green line, linear scale) and mean fragment size (blue line, linear scale) with decreasing forest area, as observed by remote sensing (dots) and simulated by the fragmentation model FRAG (solid lines) for the Americas. **b**, The number of fragments and their mean size show non-linear behaviour for all three continents (solid lines, FRAG simulations; dots, observations from remote sensing). **c**, Application of percolation theory to predict future forest fragmentation for different scenarios in the Americas using the FRAG model (see Methods for details of scenarios S1–S4).

increase in the number of forest fragments. We obtained similar trends for Africa and Asia–Australia (Fig. 3b).

We used the FRAG model to investigate how forest fragmentation would proceed in the coming decades on the basis of our suggested explanation for the observed fragmentation patterns. Under constant deforestation (at a rate of 0.51% per year in the Americas[1] without any reforestation), the number of forest fragments will greatly increase (at maximum by a factor of 33 over 50 years, scenario S1 in Fig. 3c, Extended Data Fig. 4a, e) while the mean sizes of fragments will decrease from 17 ha to 0.25 ha. To assess how reduced deforestation and reforestation may mitigate these effects, we analysed additional scenarios (Fig. 3c, Extended Data Fig. 4b–e). For example, when considering reforestation (at a rate of 0.14% per year in the Americas[1]), the deforestation rate is reduced only slightly and still leads to an increase in the number of fragments, by a factor of 28 until 2050 (scenario S2 in Fig. 3c). The same would result for a yearly reduction in deforestation rate (derived from ref. 2), for which the projected increase in fragment number is limited to a factor of 27 until 2050 (scenario S3 in Fig. 3c). Only efforts that increasingly reduce deforestation rates in combination with reforestation (for example, by 0.01% year^{-2}) would lead to increasing forest cover and decreasing fragment numbers after 20 to 30 years (scenario S4 in Fig. 3c). Such mitigation scenarios are important aspects for an assessment of forest fragmentation. Although Europe, North America and parts of Asia have experienced higher levels of reforestation than deforestation during the past few centuries[25,26], forest loss still exceeds forest gain in most tropical countries[2,26] such as Brazil, Indonesia and Cameroon.

Increases in fragment number and decreases in fragment size both have important consequences for the habitats of species[3,27,28] and for forest fragment edges[16,27]. Negative effects on biodiversity[16,27] arise from a reduction in the connectivity of fragments and the enhancement of edge effects. In addition, recent studies have outlined that increased tree mortality[16] in the edge area of forest fragments will lead to additional carbon emissions[29]. Effective policies and regulations for preventing the negative effects of fragmentation therefore require robust indicators. The development of indicators is an important challenge in environmental science, especially given the possibility that critical transitions of global ecosystems may occur as a result of human land transformations[30].

To test our results with respect to more complex spatial regimes of deforestation (the FRAG-B model), we simulated deforestation to occur with probability d_{border} at the border of already existing forest fragments and at random locations otherwise (Extended Data Fig. 5, Supplementary Video 2). In terms of criticality, the FRAG-B model showed behaviour similar to that of the classical percolation model (FRAG), for nearly the entire range of the parameter d_{border} (critical point at $p_c = 0.59$ with $\tau \sim 2$, Extended Data Fig. 6). In addition, border deforestation further widens the range of forest area in which power-law behaviour with an exponent of $\tau \sim 2$ occurs (Extended Data Fig. 7). Reforestation (for example, 0.14% per year in the Americas[1]) is already well addressed by our FRAG and FRAG-B models by lowering deforestation rates (Extended Data Fig. 8). Similarly, protection of forest (FRAG-P model) proportionally reduces the forest area that is prone to deforestation where the FRAG model still applies (simulated number of fragments scale with landscape size, Extended Data Figs 9, 10). Although local landscape structures may differ owing to local deforestation regimes, we found that the behaviour of the extended models collapses after suitable transformation in good approximation to that of the classical percolation model (FRAG). Reforestation and protecting large forest areas nevertheless have potential to mitigate the consequences of fragmentation (Extended Data Figs 4, 10).

The findings of the extended percolation models underpin the universality of fragmentation patterns close to the critical point. This is a general feature of critical phenomena: their large-scale behaviour is independent of the underlying small-scale mechanisms[7,17,21]. Even though land use appears to be complex and diverse, our outcomes emphasize that simple mechanisms are sufficient for describing forest fragmentation structures at larger scales. This finding does not exclude the possibility that the empirical patterns detected here may have alternative explanations. Our combination of spatial analyses of a high-resolution forest cover map, simulation modelling and the application of percolation theory provides a first step towards a simple explanation for the intriguing global patterns of tropical forest fragmentation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 February 2017; accepted 9 January 2018.
Published online 14 February 2018.

1. Achard, F. *et al.* Determination of tropical deforestation rates and related carbon losses from 1990 to 2010. *Glob. Change Biol.* **20,** 2540–2554 (2014).
2. Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science* **342,** 850–853 (2013).
3. Haddad, N. M. *et al.* Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci. Adv.* **1,** e1500052 (2015).
4. Chaplin-Kramer, R. *et al.* Degradation in carbon stocks near tropical forest edges. *Nat. Commun.* **6,** 10158 (2015).
5. Storch, D., Marquet, P. A. & Brown, J. H. *Scaling Biodiversity* (Cambridge Univ. Press, 2007).
6. Marquet, P. A. *et al.* Scaling and power-laws in ecological systems. *J. Exp. Biol.* **208,** 1749–1769 (2005).
7. Sornette, D. *Critical Phenomena in Natural Sciences. Chaos, Fractals, Selforganization and Disorder: Concepts and Tools* (Springer, 2006).
8. Turcotte, D. L. & Malamud, B. D. Landslides, forest fires, and earthquakes: examples of self-organized critical behavior. *Physica A* **340,** 580–589 (2004).
9. Bak, P. *How Nature Works. The Science of Self-Organized Criticality* (Copernicus, 1996).
10. Stauffer, D. & Aharony, A. *Introduction to Percolation Theory* (Taylor & Francis, 1994).
11. Christensen, K. & Moloney, N. R. *Complexity and Criticality* Vol. 1 (Imperial College Press, 2005).
12. Le Quéré, C. *et al.* Global carbon budget 2014. *Earth Syst. Sci. Data* **7,** 47–85 (2015).
13. Wright, S. J. Tropical forests in a changing environment. *Trends Ecol. Evol.* **20,** 553–560 (2005).
14. Lewis, S. L., Edwards, D. P. & Galbraith, D. Increasing human dominance of tropical forests. *Science* **349,** 827–832 (2015).
15. Houghton, R. A. The worldwide extent of land-use change. *Bioscience* **44,** 305–313 (1994).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

16. Laurance, W. F. *et al.* The fate of Amazonian forest fragments: a 32-year investigation. *Biol. Conserv.* **144,** 56–67 (2011).
17. Saberi, A. A. Recent advances in percolation theory and its applications. *Phys. Rep.* **578,** 1–32 (2015).
18. Milne, B. T. *et al.* Detection of critical densities associated with piñon-juniper woodland ecotones. *Ecology* **77,** 805–821 (1996).
19. Drossel, B. & Schwabl, F. Self-organized critical forest-fire model. *Phys. Rev. Lett.* **69,** 1629–1632 (1992).
20. Ziff, R. M. Test of scaling exponents for percolation-cluster perimeters. *Phys. Rev. Lett.* **56,** 545–548 (1986).
21. Solé, R. V. & Bascompte, J. *Self-Organization in Complex Ecosystems* Vol. 42 (Princeton Univ. Press, 2006).
22. Turner, M. G., Gardner, R. H. & O'Neill, R. V. *Landscape Ecology in Theory and Practice* (Springer, 2001).
23. Gardner, R. H., Milne, B. T., Turner, M. G. & O'Neill, R. V. Neutral models for the analysis of broadscale landscape pattern. *Landsc. Ecol.* **1,** 19–28 (1987).
24. Bascompte, J. & Solé, R. V. Habitat fragmentation and extinction thresholds in spatially explicit models. *J. Anim. Ecol.* **65,** 465–473 (1996).
25. Meyfroidt, P. & Lambin, E. F. Global forest transition: prospects for an end to deforestation. *Annu. Rev. Environ. Resour.* **36,** 343–371 (2011).
26. Rudel, T. K. *et al.* Forest transitions: towards a global understanding of land use change. *Glob. Environ. Change* **15,** 23–31 (2005).
27. Debinski, D. M. & Holt, R. D. A survey and overview of habitat fragmentation experiments. *Conserv. Biol.* **14,** 342–355 (2000).
28. Andrén, H. Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review. *Oikos* **71,** 355–366 (1994).
29. Brinck, K. *et al.* High resolution analysis of tropical forest fragmentation and its impact on the global carbon cycle. *Nat. Commun.* **8,** 14855 (2017).
30. Barnosky, A. D. *et al.* Approaching a state shift in Earth's biosphere. *Nature* **486,** 52–58 (2012).
31. R Core Team. *R: A Language and Environment for Statistical Computing*; https://www.R-project.org/ (R Foundation for Statistical Computing, 2015).
32. South, A. rworldmap: A new R package for mapping global data. *R J.* **3,** 35–43 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements The project has been supported by the Helmholtz Alliance Remote Sensing and Earth System Dynamics. A.H. and T.W. were supported by the European Research Council Advanced Grant 233066. We thank A. Hein and A. Bogdanowski for assistance, A. Hartmann for discussion, M. Dantas de Paula for data handling, and S. Paulick and F. Bohn for technical support.

Author Contributions A.H., F.T. and R.F. conceived the project. A.H., F.T. and T.W. supervised the research. S.L. and E.R. processed and analysed vegetation maps. F.T. performed statistical analysis of map observations. F.T. and M.S.M. implemented the simulation models and conducted the simulations. F.T. analysed the results and prepared figures, tables and videos. A.H., F.T., R.F., T.W. and J.G. wrote the manuscript. All authors have participated in discussion and editing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to F.T. (franziska.taubert@ufz.de).

Reviewer Information *Nature* thanks B. Barzel, B. DeVries, R. Ewers and P. Marquet for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Use and analysis of the forest cover map. The analysis of global forest fragmentation is based on Hansen's forest cover map[2] that uses Landsat observation for the year 2000 with a spatial resolution of 30 m × 30 m. A forest/non-forest classification threshold of 30% (minimum forest cover per pixel) was used in this study.

The resulting image shows non-forested and forested pixels (the original binary image files were processed as in ref. 29, ASCII format, WGS-84 projection); their connection to forest fragments is determined by their four-pixel neighbourhood (open boundary conditions). For this, an extended cluster detection algorithm was used[29]. The area of a pixel in the WGS-84 projection is calculated dependent on its geographical position[29]. As area calculation for every pixel is highly time-consuming (55 billion pixels for land surface in the tropics), we pre-calculated the area size of pixels along 256 latitudes. No statistical methods were used to predetermine sample size.

Forest fragmentation model FRAG. For simulation of fragmentation dynamics we used a landscape with C_{max} cells of size s (30 m × 30 m). Cells can have two states (forested or deforested). We start the simulation with a fully forested area (all cells are in the 'forest' state, total forest area is A_{max} (landscape area), $A_{max} = C_{max}\, s$). In each step (here, one year) some forest area is cleared assuming a constant deforestation rate d (% per year). That is, a certain number of forest cells ($C_{max}\, d$) are randomly selected and assigned to the state 'deforested' in each year.

For simplicity, cells do not regenerate back to forest in the FRAG model, so forest area is successively reduced over time until the entire forest area is cleared. However, deforestation rates used in the FRAG model can also be interpreted as net deforestation rates that result from reforestation and gross deforestation occurring at random sites. In detail, with d being the yearly gross deforestation rate (for example, 0.51% per year derived from ref. 1 for the tropical Americas) and r being the yearly reforestation rate (for example, 0.14% per year derived from ref. 1 for the tropical Americas), we obtained a yearly net deforestation rate of $d_{net} = d - r$. Including random reforestation into our simulations therefore shows that the effect of reforestation is already covered by the FRAG model (Extended Data Fig. 8a). If reforestation occurs exclusively at the border of forest fragments, small deviations can be observed (median deviations show 5% lower fragment numbers than for random reforestation).

We analysed the results of the fragmentation model using the same methods as for the analysis of the high-resolution forest cover map in terms of remaining forest area, fragment numbers and their mean size as well as the fragment size distribution (see 'Analysis and statistics of tropical forest fragmentation'). Dynamics of fragment numbers scale with landscape size C_{max} (total number of cells, Extended Data Fig. 9a). Normalized fragment numbers can therefore be calculated by dividing the absolute number of fragments by the total number of cells (C_{max}).

The critical point p_c is determined as the remaining forest area relative to landscape area A_{max} after which the spanning cluster no longer exists. The spanning cluster is defined as any large cluster that expands from one border of the landscape to an opposite side border (either north to south or east to west). Note that this simple fragmentation model corresponds to a dynamic version of the classical percolation model[10,11,22,23].

Simulation of forest fragmentation at the continental scale. Potential forest areas before forest clearing have been estimated for each continent using a map of vegetation biomes[33] selecting for classified tropical forest area (only biomes of tropical and subtropical woodland). Potential areas have been calculated by remapping vegetation cover of gridded maps (0.5° × 0.5°) to an equidistant map (1 km × 1 km) using climate data operators[34].

We simulated forest fragmentation dynamics by using, for each continent, the estimated potential forest cover (cell size s = 30 m × 30 m) and deforestation rates derived from literature (Table 1). Extents of simulated continental areas for the fragmentation model were 1,377 million ha for the Americas, 806 million ha for Africa and 770 million ha for Asia–Australia (corresponding to C_{max} values of 15,300,205,636 cells for the Americas, 8,955,593,956 cells for Africa and 8,555,695,009 cells for Asia–Australia). To determine the current forest area (in %) in Fig. 3a, we divided the observed forest areas (from remote sensing, Table 1) by the estimated potential forest areas. Gross deforestation rates (Table 1) and reforestation rates were derived from ref. 1 by dividing area change estimates of annual gross deforestation (2000–2010) and annual forest regrowth (2000–2010) by area estimates of forest cover (2000) for dry and humid tropical forests. It should be noted that potential forest area estimates can include uncertainties that can affect the determined fraction of current forest area (in %) for the forest cover map.

Simulation scenarios of future forest fragmentation patterns. We simulated four different scenarios (S1 to S4) to project future development of forest fragmentation (until 2050 for the Americas, using the FRAG model, Extended Data Fig. 4):

Scenario S1: Net deforestation rate is constant at d_{net} = 0.51% per year (ref. 1). This scenario assumes no reforestation.

Scenario S2: Deforestation of d = 0.51% per year (ref. 1) is counterbalanced by reforestation of r = 0.14% per year (ref. 1). This results in a net deforestation rate of d_{net} = 0.37% per year.

Scenario S3: Deforestation rate of 0.37% per year (ref. 1) is reduced by 0.0012% year^{-2}. This results in a net deforestation rate of $d_{net} = a - bt$, with t being the number of years since the critical point, a = 0.37% per year and b = 0.0012% year^{-2}. This scenario for the tropical Americas was based on a 12-year observation of forest gain and forest loss (ref. 2, version v1.3). The derived absolute values of net forest area change were related to our simulated landscape area (A_{max}) and fitted by linear regression (slope = 0.0012, R^2 = 0.013).

Scenario S4: Deforestation of 0.37% per year (ref. 1) is reduced by 0.01% year^{-2}. This results in a net deforestation rate of $d_{net} = a - bt$ (with a = 0.37% per year and b = 0.01% year^{-2}) with a turning point at which forest recovery exceeds deforestation. This scenario is similar to S3, but with a stronger decreasing trend of yearly deforestation rates.

Projections of future fragmentation for the year 2050 started from the critical point of percolation using the FRAG model (as a simplification). A time step corresponds to a constant reduction of forest area per year. Projected values (of fragment numbers and mean size in 2050) for different deforestation scenarios (Fig. 3c) were compared to observational values from remote sensing (Table 1, factors shown in Extended Data Fig. 4). The fragmentation dynamics (in terms of deforestation rate, forest cover, fragment numbers and mean fragment size) for each scenario are shown in Extended Data Fig. 4 (absolute values can be found in Extended Data Fig. 4e).

Forest fragmentation model FRAG-B. In addition, we analysed a second version of the fragmentation model in order to test for more complex deforestation patterns. Here, a probability d_{border} of clearing only forested cells at the border of a forest fragment has been introduced. When d_{border} = 0, forested cells are cleared randomly (FRAG model version). When d_{border} = 0.5, 50% of the deforestation is restricted to fragment borders while the remaining 50% is still deforested randomly throughout the area. When d_{border} = 1, only cells at the border of forest fragments are cleared (with random site selection for deforestation in the first time step). We analysed the result of the FRAG-B model similar to that of the FRAG model (see 'Analysis and statistics of tropical forest fragmentation'). Dynamics of fragment number also scale with landscape size C_{max} (Extended Data Fig. 9b). As such, normalized fragment numbers can again be calculated by dividing the absolute number of fragments by the total number of cells (C_{max}).

Gross deforestation rates used in the FRAG-B model can again be interpreted as net deforestation rates that result from reforestation at random sites together with border gross deforestation (similar to the FRAG model, Extended Data Fig. 8b). In this case, the probability that deforestation occurs at the border of forest fragments (d_{border}) becomes approximately $d_{border,net} = (d_{border}\, d - r)/(d - r)$, with d being the yearly gross deforestation rate (for example, 0.51% per year in the Americas[1]) and r being the yearly reforestation rate (for example, 0.14% per year in the Americas[1]). Small deviations can be attributed to the dynamic change of border area of fragments during the simulation. By contrast, reforestation is fully captured by our FRAG-B model if reforestation occurs with the same probability d_{border} at the border of forest fragments as does deforestation.

In order to test for a full range of border deforestation, we simulated a forest landscape ($C_{max} = 10^6$ cells) and varied d_{border} from 0 to 1 in steps of 0.1. Between d_{border} = 0.9 and d_{border} = 1.0 we tested model behaviour for detailed values: d_{border} = 0.925, 0.95, 0.975, 0.98, 0.985, 0.99, and 0.995 (because critical points and power-law exponents of fragment sizes drop between d_{border} = 0.9 and d_{border} = 1.0). We focused our analysis on the critical points and the fitted power-law exponents of fragment size distributions for each simulation (Extended Data Fig. 6) with the following sample sizes of fragments: n = 29,788 (d_{border} = 0), n = 30,195 (d_{border} = 0.1), n = 29,020 (d_{border} = 0.2), n = 28,131 (d_{border} = 0.3), n = 31,117 (d_{border} = 0.4), n = 30,506 (d_{border} = 0.5), n = 30,733 (d_{border} = 0.6), n = 34,182 (d_{border} = 0.7), n = 35,325 (d_{border} = 0.8), n = 43,045 (d_{border} = 0.9), n = 44,997 (d_{border} = 0.925), n = 45,964 (d_{border} = 0.95), n = 45,577 (d_{border} = 0.975), n = 44,415 (d_{border} = 0.98), n = 45,692 (d_{border} = 0.985), n = 43,806 (d_{border} = 0.99), n = 38,875 (d_{border} = 0.995), n = 1,278 (d_{border} = 1.0).

Forest fragmentation model FRAG-P. A third version of the fragmentation model includes protected forest areas. Here, a fraction $f_{protected}$ of landscape area (in total $f_{protected}\, A_{max}$) is preserved against deforestation. When $f_{protected}$ = 0, forest area is cleared randomly (FRAG model). When $f_{protected}$ = 0.5, 50% of forest area is protected while the remaining 50% is still deforested randomly. When $f_{protected}$ = 1, no forest area is cleared. The protected forest area is distributed among rectangular forest patches (10,000 ha in size) that are randomly distributed across the landscape. In order to reach the total protected area exactly ($f_{protected}\, A_{max}$), a few forest patches of different sizes are also assigned for protection (for example, if protected forest patches overlap or are placed too close to the landscape boundaries).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Conceptually, protecting large forest fragments means preserving certain size classes in the fragment size distribution (for example, in our FRAG-P model the sizes from 10^4 ha to 10^5 ha). Therefore, protection of one large fragment is equivalent to excluding this fragment from the study area prone to deforestation (for example, similar to the FRAG model). Because the normalized number of fragments (normalization means dividing by the total number of cells C_{max}) is independent of landscape size, we obtain a behaviour that is captured by the FRAG model for smaller landscape areas as well (Extended Data Fig. 9). This property reduces the extrapolated number of forest fragments proportionally by $(1 - f_{protected})$ (Extended Data Fig. 10).

In order to test for a range of protected forest area, we simulated different fractions of protected areas ($f_{protected} = 0.1$ and 0.5, for landscape sizes compared to the tropical Americas; that is, $C_{max} = 15{,}300{,}205{,}636$ cells). We focused our analysis on the number of forest fragments (Extended Data Fig. 10).

Analysis and statistics of tropical forest fragmentation. Forest fragments were detected by applying an extended cluster detection algorithm[29]. This algorithm labels each single fragment of cells on the basis of the direct four neighbours per cell (see 'Use and analysis of the forest cover map'). The mean fragment size of a landscape is calculated by summing up their sizes divided by the number of fragments (unweighted mean including the spanning cluster).

We calculated the fragment size distribution, that is the number of fragments per fragment size class. In order to describe its form, we applied multinomial maximum likelihood estimation of logarithmically binned data to fragment sizes published in ref. 35 (using the MATLAB package provided online). We fitted the frequency of fragment sizes to a power-law distribution to determine the scaling exponent ($N(f) \sim f^{-\tau}$, where $N(f)$ is the number of fragments of size f and τ is the scaling exponent). For logarithmic binning, we used size classes (in ha) of $[10^1, 10^2)$, $[10^2, 10^3)$, … $[10^8, 10^9)$. We excluded fragments smaller than 10 ha for the fit. Standard errors[35] for each fitted exponent were calculated by bootstrapping (1,000 repetitions). Goodness of fit was evaluated by using Pearson's correlation coefficient (R) between predicted and empirical cumulative distributions (both logarithmic; using raw data including also empirical and simulated fragments smaller than 10 ha). Similar to fragment sizes, we calculated perimeters (edge length) of detected forest fragments for analysing and fitting the fragment perimeter distribution ($N(l) \sim l^{-\kappa}$, where $N(l)$ is the frequency of fragments with perimeter l and κ is the scaling exponent). For logarithmic binning, we used perimeter classes $[10^3, 10^4)$, $[10^4, 10^5)$, … $[10^9, 10^{10})$ in m and excluded fragment perimeters smaller than 1,000 m for the fit (equivalent to a fragment size of approximately 10 ha). Standard errors[35] were again derived for each fitted exponent by bootstrapping (1,000 repetitions). Goodness of fit was similarly evaluated by using Pearson's correlation coefficient (R) between predicted and empirical cumulative distributions (both logarithmic; using raw data including also empirical and simulated fragment perimeters smaller than 1,000 m).

We further determined the fractal dimension d_f on the basis of subsequently grouping 30 m × 30 m cells of the vegetation map to boxes of $b_f = 2^1, 2^2, 2^3, \ldots, 2^8$ cells (box counting method[36,37]). For each box, we determined how many cells include forest. When at least one cell per box is occupied by forest, we mark the entire box as forested. By this method, a map is created that has coarser resolution than the original map (with $s = 30\,m \times 30\,m$ resolution). The fractal dimension is then calculated by relating the inverse of b_f to the number of boxes including forest on the landscape of coarser resolution. This relationship follows a power law, the scaling exponent of which represents the fractal dimension. We fitted this relationship on log–log axes by using ordinary least squares regression and calculated the corresponding coefficient of determination R^2. We calculated the fractal dimension d_f for the high-resolution vegetation map (Table 1, with R^2 values of approximately 1 for each continent) and compared them to the predicted dimension $d_f = 1.89$ of percolation theory[11]. It should be noted that deviations in the exponent τ of the fragment size distribution trace back to the finite area of the continents, whereas the calculation of fractal dimension is not affected by the finite area.

The large extent of our simulated forest landscapes shows low variations in results for ten replicated simulations. We calculated the coefficient of variation CV (standard deviation σ divided by mean value μ, $CV = \sigma/\mu$) from ten repeated simulations for each time step. Finally, we determined the median value, CV_{median}, as an aggregated statistic for all time steps for the number and mean size of fragments. For the fragment size distribution, we calculated the CV only at the critical point for each class of fragment sizes and determined the median CV_{median} for all size classes (details of size classes are provided earlier in this section). For fragment number and mean size, median CV_{median} values were below 0.00004 for all three continents (using the FRAG model, CV_{median} decreased in general during the simulation time). Median CV_{median} among size classes of the fragment size distribution (at the critical point) were less than 0.0062 (CV_{median} decreased in general with smaller size classes). Similar results were obtained for the FRAG-B (0.00004 and 0.0036 for $d_{border} = 0.9$, landscape area of the Americas) and FRAG-P models (0.00003 and 0.448 for $f_{protected} = 0.1$, landscape area of the Americas). Because stochasticity did not translate into visible effects, we visualized only one simulation run in Fig. 2, Fig. 3 and Extended Data Figs 3–10 and provided absolute values of one simulation run in Extended Data Fig. 4e.

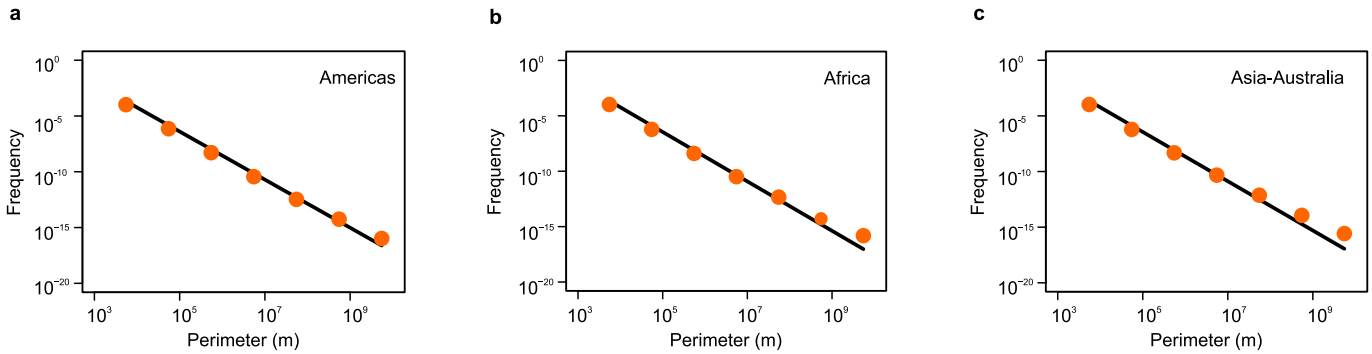
Visualization of fragmentation patterns. We gathered snapshots of fragment size distributions and spatial patterns for each simulated year in our FRAG and FRAG-B models for the Americas, as MPG files (Supplementary Videos 1, 2) and figures (Fig. 2, Extended Data Figs 3, 5). For the fragment size distribution, green bars show the simulation results (FRAG or FRAG-B models) and the black line shows the observed size distribution (from remote sensing) for the tropical forest in the Americas. We used logarithmic binning of fragment size classes $[10^1, 10^2)$, $[10^2, 10^3)$, … $[10^9, 10^{10})$ in ha. For graphical purposes only, we additionally visualized size classes $[10^{-1}, 10^0)$ and $[10^0, 10^1)$. The remaining forest area (p) is provided within the bar plot of the videos. For the map shown (a selected subarea of 900 ha was used for Extended Data Figs 3 and 5 as well as for Supplementary Videos 1 and 2; a selected subarea of 225 ha was used for Fig. 2 for graphical purposes), each fragment is coloured according to its size. Six size classes (x, in ha) were chosen with the following colour codes: $x < 0.4$ (light blue), $0.4 \le x < 2$ (yellow), $2 \le x < 9$ (light magenta), $9 \le x < 42$ (light green), $42 \le x < 195$ (dark magenta), $x \ge 195$ (dark green). White cells denote cleared sites.

Code availability. The extended cluster detection software and the simulation models are available upon request from the corresponding author.

Data availability. The observed fragment-size and perimeter-distribution data, as well as the data of the simulation studies, are available upon request from the corresponding author.

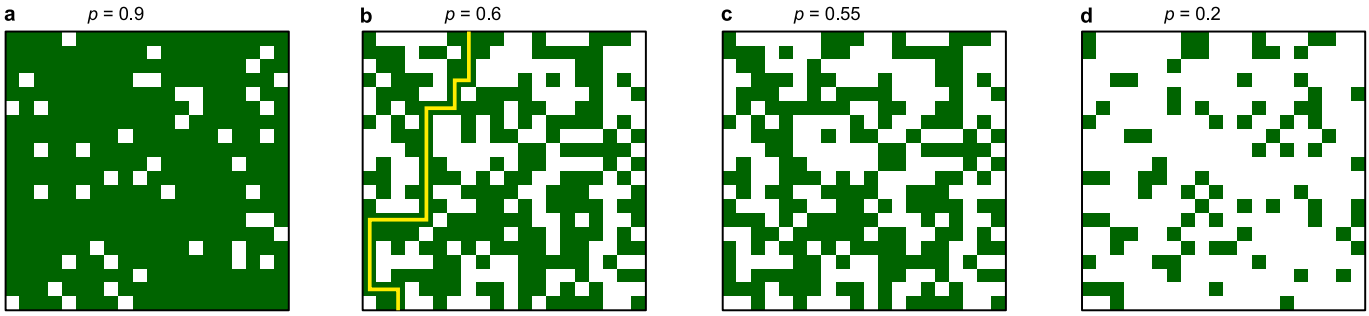
33. Ostberg, S., Schaphoff, S., Lucht, W. & Gerten, D. Three centuries of dual pressure from land use and climate change on the biosphere. *Environ. Res. Lett.* **10,** 044011 (2015).
34. CDO 2015: Climate Data Operators. *The Max Planck Institute for Meteorology*, http://www.mpimet.mpg.de/cdo (2017).
35. Virkar, Y. & Clauset, A. Power-law distributions in binned empirical data. *Ann. Appl. Stat.* **8,** 89–119 (2014).
36. Sugihara, G. & May, R. M. Applications of fractals in ecology. *Trends Ecol. Evol.* **5,** 79–86 (1990).
37. Bisoi, A. K. & Mishra, J. On calculation of fractal dimension of images. *Pattern Recognit. Lett.* **22,** 631–637 (2001).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



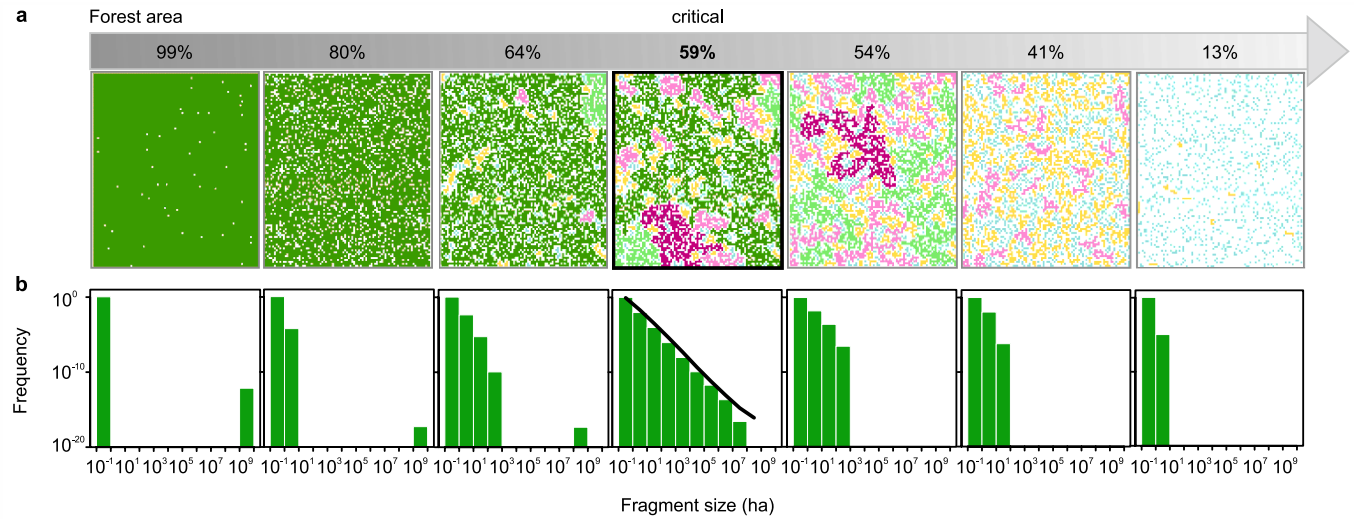
Extended Data Figure 1 | Continental-scale fragment perimeter distribution of tropical and subtropical forest. a–c, Observed fragment perimeter distributions (orange dots; fragment perimeters >1,000 m) for the Americas (**a**, $n = 55.5$ million fragments), Africa (**b**, $n = 44.8$ million fragments) and Asia–Australia (**c**, $n = 30.5$ million fragments); solid black lines show the fitted power laws with exponent κ. Forest fragments were estimated from Hansen's forest cover map[2] (see Methods for details).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



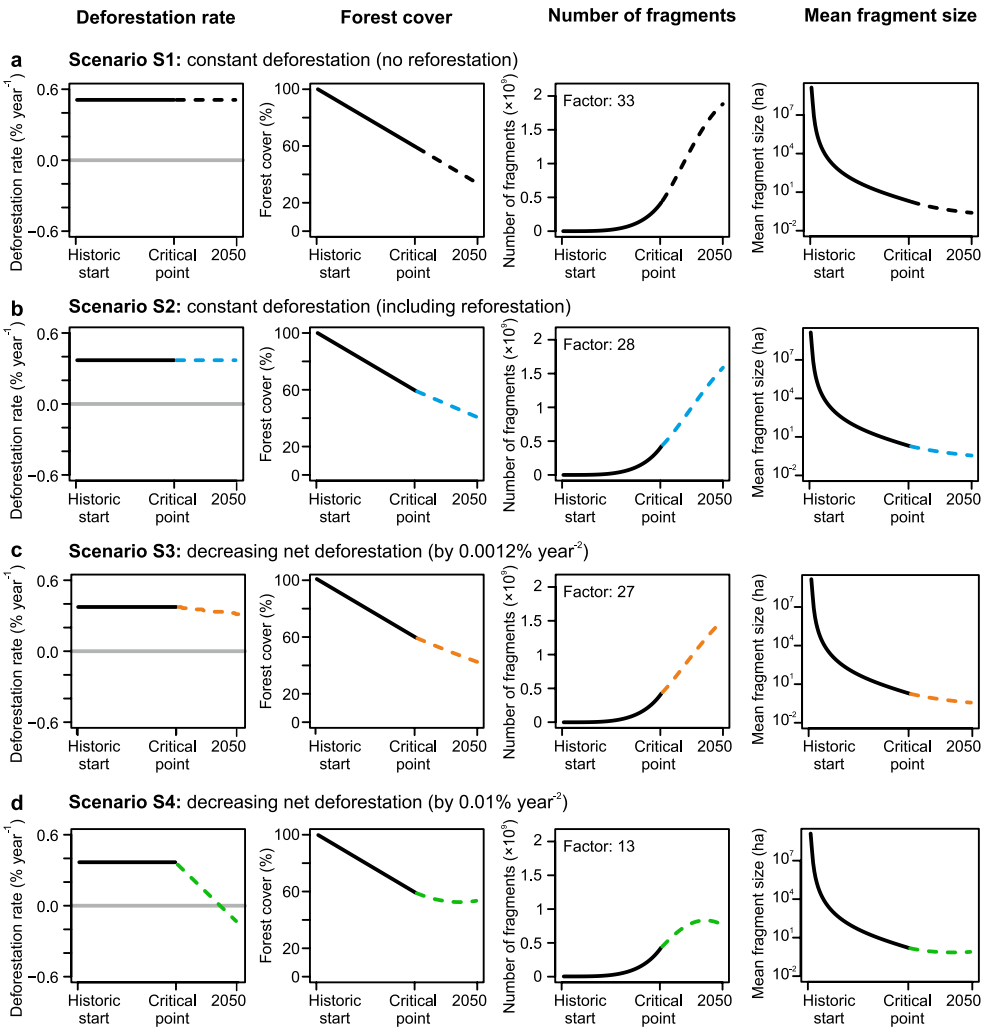
Extended Data Figure 2 | Illustration of the concept of classical percolation theory. a, A landscape occupied by 90% randomly and independently distributed forest cells (green, $p = 0.9$) is dominated by one large cluster. **b**, For values of p larger than the percolation threshold ($p = 0.6 > 0.59$) there exists a continuous path from two opposite-side borders (yellow path within the spanning cluster). **c**, Below the percolation threshold (here, $p = 0.55 < 0.59$) larger clusters emerge but no spanning cluster can be detected. **d**, A landscape occupied by 20% forest shows small unconnected clusters. Landscape size is 20×20 cells.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 3 | Dynamics of tropical forest fragmentation in the Americas (FRAG model). a, Spatial patterns of fragments for different snapshots in time, and **b**, fragment size distributions (green bars indicate values from the FRAG model, solid black line indicates observation from remote sensing; fragment sizes ≥ 0.1 ha). The critical phase at which the spanning fragment disappears is indicated as 'critical'. For each phase, a map of a selected subarea of 900 ha is shown (from the FRAG model, cleared sites are white and colours indicate fragment size; see Methods for details). See also Supplementary Video 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



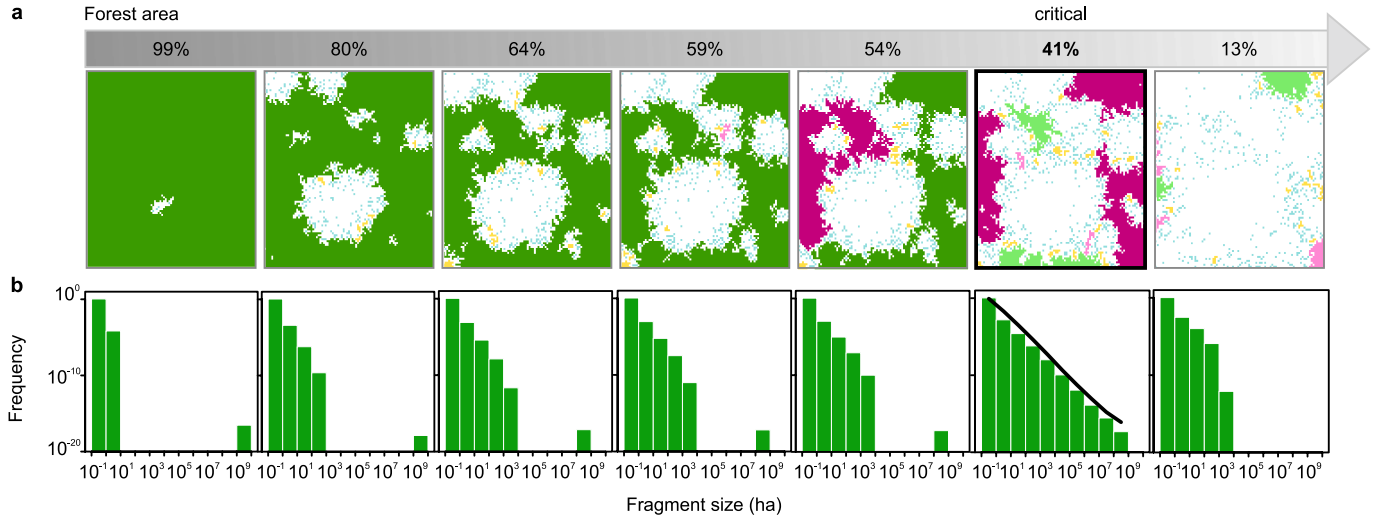
e Results of scenarios

	Observation*	Scenario S1†	Scenario S2†	Scenario S3†	Scenario S4†
Number of fragments (×10^6)	55.5	1,880.4	1,587.4	1,507.2	772.1
Mean fragment size (ha)	17	0.25	0.35	0.39	0.95

* year 2000
† for year 2050

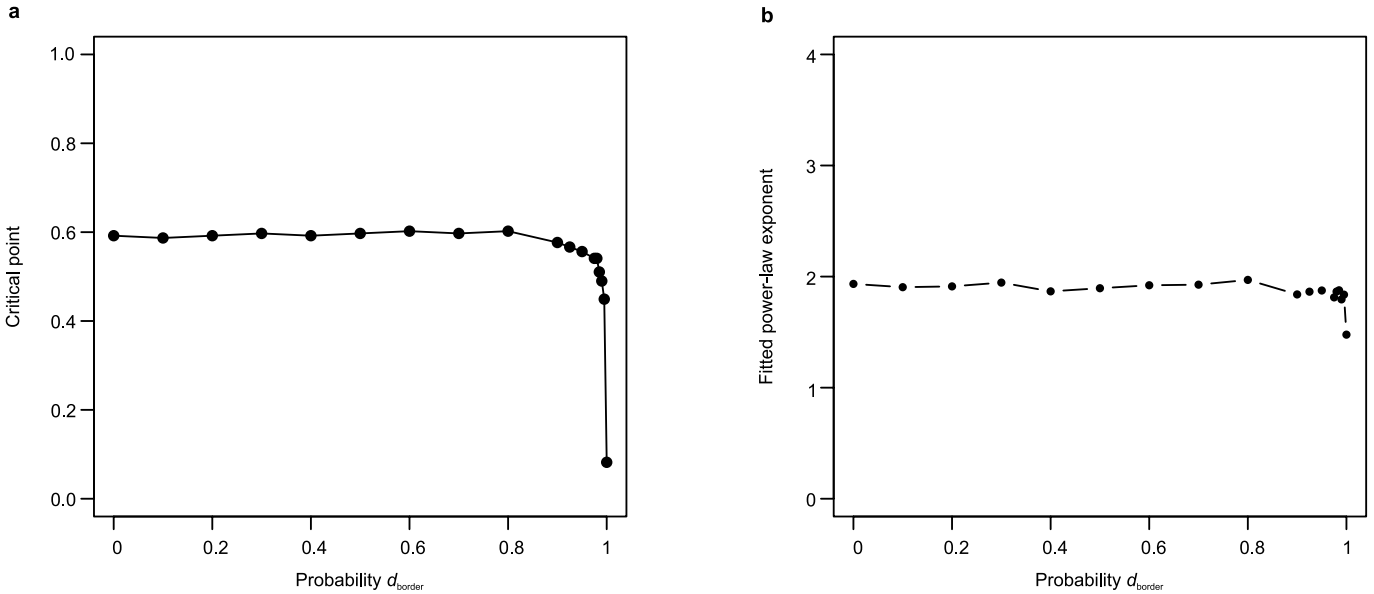
Extended Data Figure 4 | Dynamics of fragmentation scenarios projected until 2050. Dynamics of deforestation rate (percentage per year related to landscape size), forest cover (%), number of fragments and mean fragment size (in ha) in America projected until 2050 using the FRAG model. **a**, Scenario S1 assumes constant deforestation without reforestation. **b**, Scenario S2 considers added reforestation. **c**, Scenario S3 assumes that deforestation decreases yearly. **d**, Scenario S4 considers a stronger yearly reduction in deforestation, leading to a turning point with net reforestation. See Methods for details. **e**, Comparison of observed (Table 1) and projected fragment numbers (rounded) and mean fragment size.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



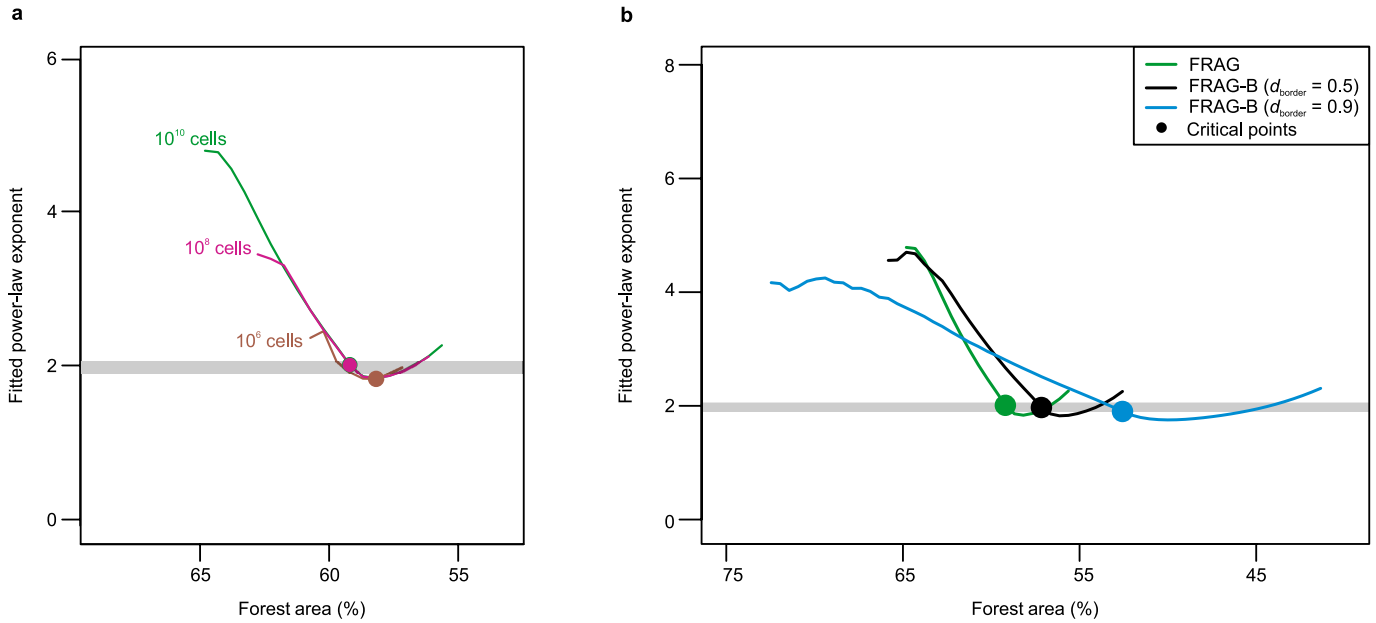
Extended Data Figure 5 | Dynamics of tropical forest fragmentation in the Americas (FRAG-B model). a, Spatial patterns of fragments for different snapshots in time, and **b**, fragment size distributions (green bars indicate values from the FRAG-B model, $d_{border} = 0.995$; solid black line denotes observation from remote sensing; fragment sizes ≥ 0.1 ha). The critical phase at which the spanning fragment disappears is indicated as 'critical'. For each phase, a map of a selected subarea of 900 ha is shown (from the FRAG-B model, cleared sites are white and colours indicate fragment size; see Methods for details). See also Supplementary Video 2.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



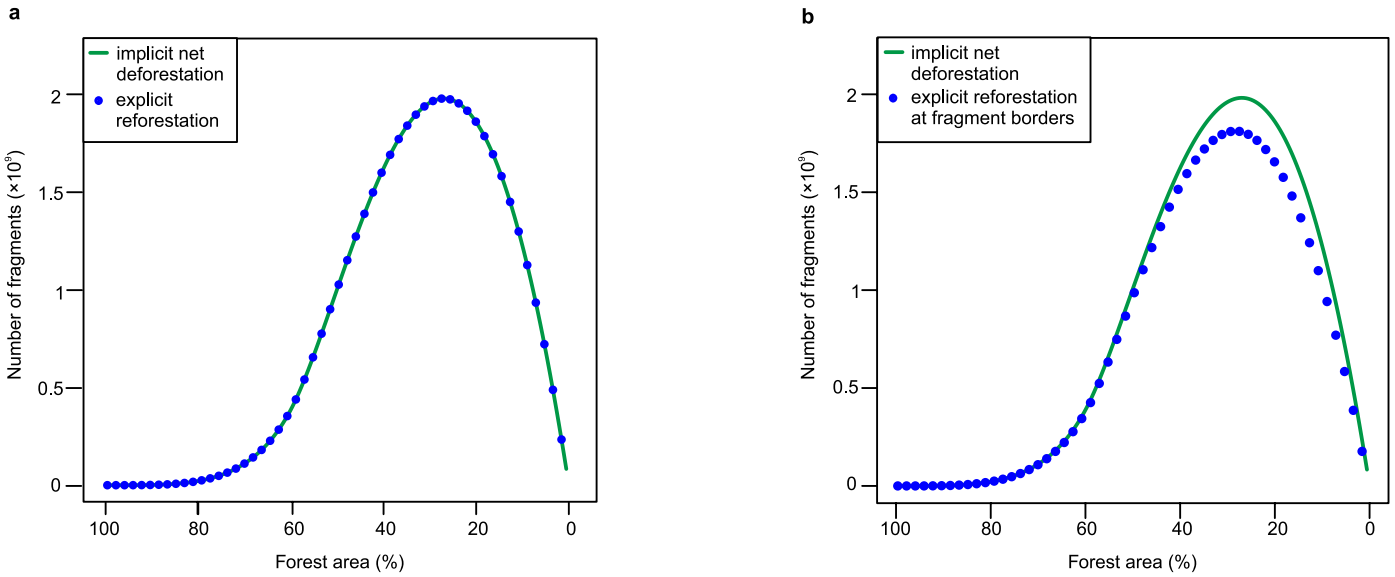
Extended Data Figure 6 | The effect of border deforestation on fragmentation dynamics. a, Critical points and **b**, fitted power-law exponents of fragment sizes for different probabilities (d_{border}) of deforestation restricted to the border of forest fragments (FRAG-B model for landscape size of $C_{max} = 10^6$ cells). Each point in **b** represents one power-law fit based on $n = 29{,}788$ ($d_{border} = 0$) to $n = 1{,}278$ ($d_{border} = 1.0$) fragments (see Methods for detailed values). Calculated R^2 values of fits are approximately 1 for the entire range of d_{border} probabilities (see Methods).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



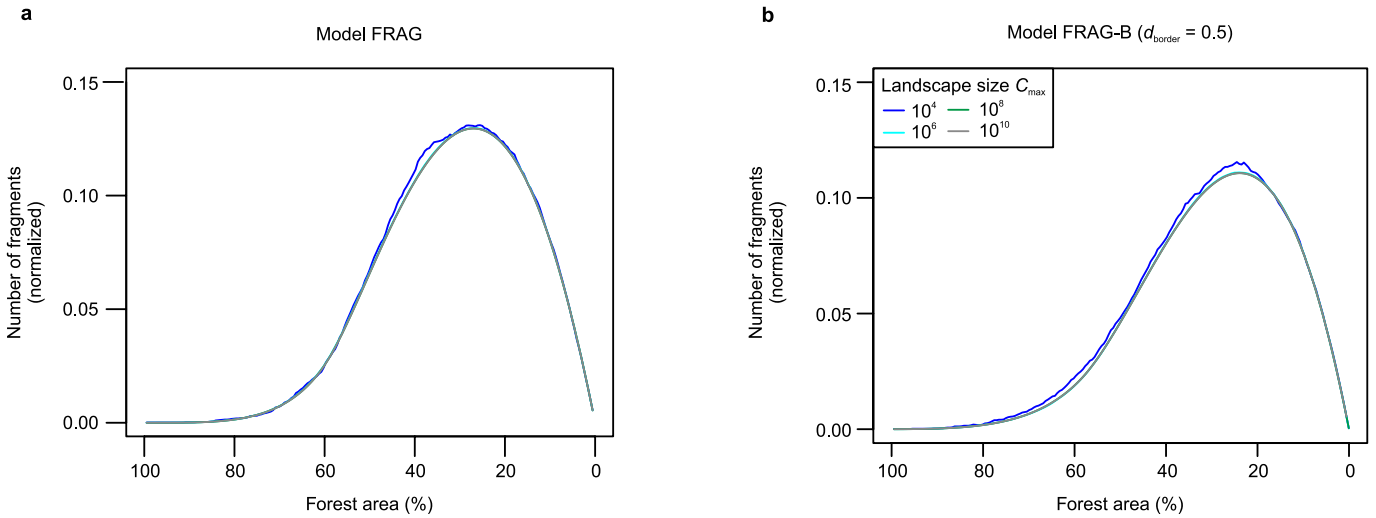
Extended Data Figure 7 | Power-law exponents fitted to simulated fragment size distributions. a, For different landscape sizes, with C_{max} based on the FRAG model. **b**, Comparison between random (FRAG model) and border deforestation (FRAG-B model, $d_{border} = 0.5$ and $d_{border} = 0.9$). In **b**, the landscape size was $C_{max} = 10^{10}$ cells. In both **a** and **b** we show only results for fits with correlation coefficients R^2 of ≥ 0.9. The grey horizontal line shows the exponent predicted by percolation theory.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



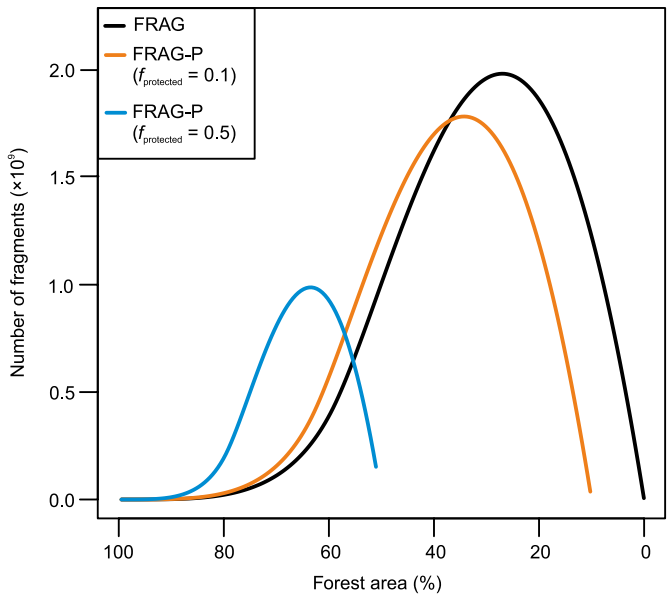
Extended Data Figure 8 | Dynamics of fragment numbers comparing implicit and explicit modelling of reforestation. a, We assume a gross deforestation rate of $d = 0.51\%$ per year and a reforestation rate of $r = 0.14\%$ per year for the Americas[1] (with random and independent selection of sites for deforestation and reforestation, blue dots). This model version is equivalent to the original FRAG model assuming a net deforestation rate of $d_{net} = 0.37\%$ per year (green line). **b**, The same scenario (green line as in **a**), but now explicit reforestation occurs exclusively at the border of forest fragments (blue dots).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 9 | Scaling of fragmentation dynamics. Dynamics of forest fragment numbers (normalized by landscape size C_{max}) in America using **a**, the FRAG model and **b**, the FRAG-B model ($d_{border} = 0.5$) for different landscape sizes. The pattern is independent of landscape size.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | Dynamics of the number of forest fragments, without (FRAG model) and with (FRAG-P model) consideration of protected forest areas. Simulations with the FRAG-P model account for 10% ($f_{protected} = 0.1$, orange) and 50% ($f_{protected} = 0.5$, blue) of the landscape area to be protected while the remaining forest area is prone to deforestation (FRAG model). Forest areas affected by deforestation in both models were simulated using a deforestation rate of $d = 0.51\%$ per year in the Americas[1].

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature25742

Tissue stiffening coordinates morphogenesis by triggering collective cell migration *in vivo*

Elias H. Barriga[1,2], Kristian Franze[3], Guillaume Charras[1,2] & Roberto Mayor[1]

Collective cell migration is essential for morphogenesis, tissue remodelling and cancer invasion[1,2]. *In vivo*, groups of cells move in an orchestrated way through tissues. This movement involves mechanical as well as molecular interactions between cells and their environment. While the role of molecular signals in collective cell migration is comparatively well understood[1,2], how tissue mechanics influence collective cell migration *in vivo* remains unknown. Here we investigated the importance of mechanical cues in the collective migration of the *Xenopus laevis* neural crest cells, an embryonic cell population whose migratory behaviour has been likened to cancer invasion[3]. We found that, during morphogenesis, the head mesoderm underlying the cephalic neural crest stiffens. This stiffening initiates an epithelial-to-mesenchymal transition in neural crest cells and triggers their collective migration. To detect changes in their mechanical environment, neural crest cells use mechanosensation mediated by the integrin–vinculin–talin complex. By performing mechanical and molecular manipulations, we show that mesoderm stiffening is necessary and sufficient to trigger neural crest migration. Finally, we demonstrate that convergent extension of the mesoderm, which starts during gastrulation, leads to increased mesoderm stiffness by increasing the cell density underneath the neural crest. These results show that convergent extension of the mesoderm has a role as a mechanical coordinator of morphogenesis, and reveal a link between two apparently unconnected processes—gastrulation and neural crest migration—via changes in tissue mechanics. Overall, we demonstrate that changes in substrate stiffness can trigger collective cell migration by promoting epithelial-to-mesenchymal transition *in vivo*. More broadly, our results raise the idea that tissue mechanics combines with molecular effectors to coordinate morphogenesis[4].

The neural crest migrates as a mesenchymal cell population, and key transcription factors (for example, those in the Snail, Twist and Zeb families) are sufficient to promote epithelial-to-mesenchymal transition (EMT) *in vitro*[3,5]. During development, these factors are expressed well in advance of the onset of neural crest migration[3,6]; however it is unclear whether they are sufficient to trigger cell migration or whether additional extrinsic factors are also needed. To test whether the onset of neural crest migration is determined by extrinsic factors, we used heterochronic tissue grafts to expose non- and pre-migratory neural crest to embryonic environments at different developmental stages. Here we define non- and pre-migratory neural crest as those cells that are not migrating (stage 13) or are about to start migrating (stage 20), respectively[7] (Fig. 1a). When grafted into a host at a migratory stage, non-migratory neural crest migrated, whereas migration was impaired if pre-migratory neural crest was grafted into a non-migratory host (Fig. 1b–e; graft controls in Extended Data Fig. 1a–d). Furthermore, when cultured *ex vivo* and exposed to the neural crest chemoattractant Sdf-1 (also known as Cxcl12), which controls directional neural crest migration *in vivo*[8], non- and pre-migratory neural crests were identically motile (Fig. 1f; Supplementary Video 1), indicating that the cells are intrinsically capable of migrating irrespective of their embryonic stage. Together, these observations suggest that changes in environmental factors regulate the onset of neural crest migration.

One possible source of environmental changes is modification of the extracellular matrix (ECM). However, we did not observe changes in fibronectin, the principal component of *Xenopus* cranial neural crest ECM[9], between non- and pre- migratory stages (Extended Data Fig. 2a–d). As both EMT and cell migration have been shown to be influenced by the mechanical properties of the cellular environment *in vitro*[10,11], we next characterized the tissue stiffness of developing *Xenopus* embryos using iAFM, an *in vivo* atomic force microscopy (AFM) approach[12]. As the cephalic neural crest uses head mesoderm as a substrate for migration, we directly measured the apparent elastic modulus of the mesoderm by removing the superficial epidermis (Fig. 1g; AFM controls in Extended Data Fig. 2g–j). When compared at non- and pre-migratory embryonic stages, the stiffness of the mesoderm in front of the neural crest gradually and substantially increased over time (Fig. 1h). Removal of fibronectin had no effect on elastic modulus (Extended Data Fig. 2e, f), confirming that fibronectin does not contribute to mesodermal stiffness[13]. We found a strong correlation between mesodermal stiffening and the onset of neural crest migration ($R = 0.82$, $n = 16$ embryos), suggesting that tissue stiffening may trigger neural crest collective cell migration (CCM) *in vivo*.

To test whether these changes in *in vivo* substrate stiffness are sufficient to trigger neural crest CCM, we cultured neural crests on fibronectin-coated hydrogels with stiffness values similar to those found in non- and pre-migratory mesoderm (Extended Data Fig. 2k–m). Remarkably, pre-migratory neural crest migrated towards Sdf-1 when explanted onto stiff substrates but not when explanted onto soft substrates (Fig. 1i and Supplementary Video 2). High-resolution imaging revealed that clusters or individual neural crest cells explanted onto stiff substrates formed larger protrusions than those explanted onto soft substrates (Extended Data Fig. 3a–c and Supplementary Video 3). Neural crest clusters, but not single cells, displayed directional motion towards Sdf-1 on stiff substrates (Extended Data Fig. 3d). These observations suggest that substrate stiffness can be sensed at the single cell level; however, directed motion is an emergent property arising from cell–cell interactions.

Furthermore, neural crest cultured on stiff, but not soft, substrates tended to disperse (Extended Data Fig. 3e–h and Supplementary Video 4), a characteristic of EMT[14]. Consequently, we found that a stiff substrate reduced the levels of the epithelial marker E-cadherin[2,14], whereas the expression of the mesenchymal marker N-cadherin[2,8,14] was increased (Extended Data Fig. 3i, j). Hence, environmental stiffening of the mesoderm may prime neural crest CCM by triggering EMT.

To confirm that the observed increase in mesodermal stiffness is important for neural crest migration *in vivo*, we next perturbed mesoderm stiffness using mechanical and molecular manipulations. Most developmental systems are under tension[15,16], which can

[1]Department of Cell and Developmental Biology, University College London, Gower Street, London WC1E 6BT, UK. [2]London Centre for Nanotechnology, University College London, Gower Street, London WC1H 0AH, UK. [3]Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

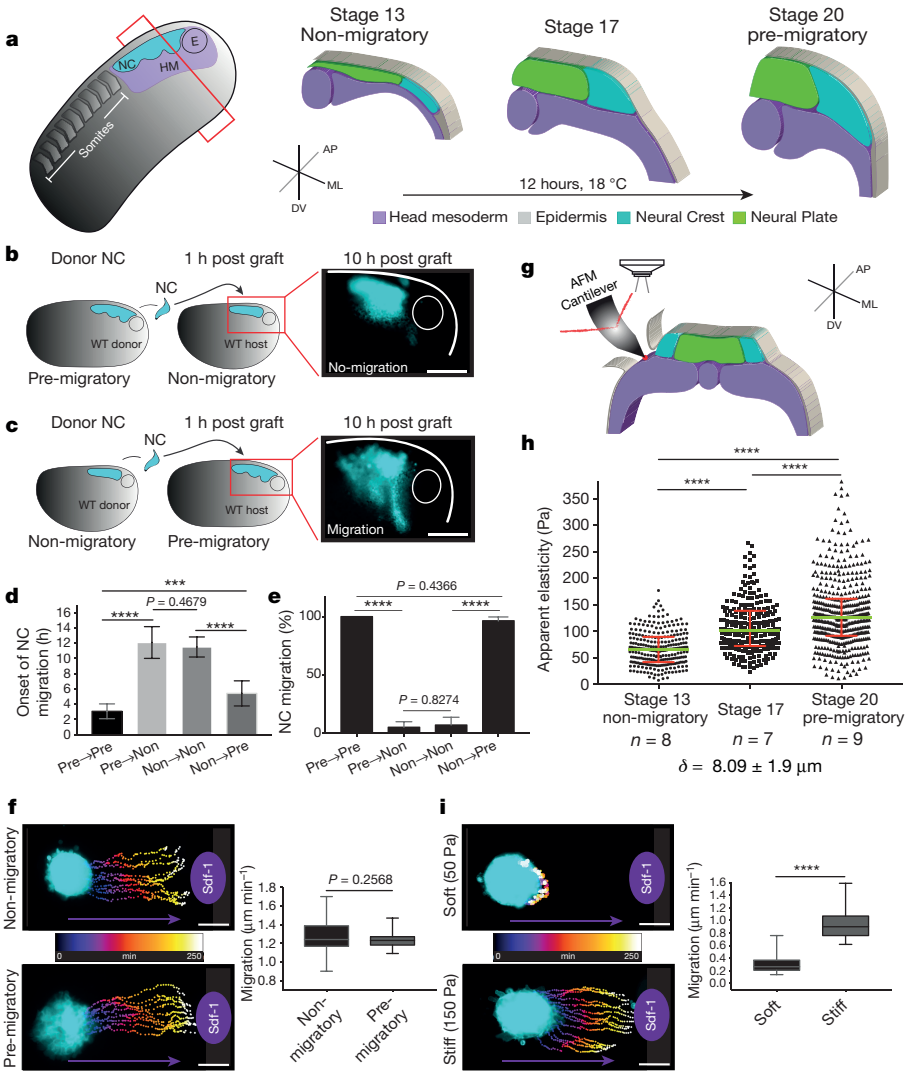


Figure 1 | Changes in environmental stiffness are required for neural crest CCM. a, Schematic cross-sections of neural crest development (AP, anteroposterior; ML, mediolateral; DV dorsoventral; NC, neural crest; HM, head mesoderm; E, eye). Cephalic neural crest originates from ectoderm at the border of the neural plate and undergoes EMT before migrating by using head mesoderm as substrate. **b**–**e**, Heterochronic grafts. **b**, Labelled pre-migratory neural crest (cyan) grafted into non-migratory hosts, a representative example of 'not-migrating' neural crest at 10 h post-graft is shown. WT, wild-type. **c**, Non-migratory neural crest (cyan) grafted into pre-migratory hosts, a representative example of 'migrating' neural crest at 10 h post-graft is shown. Scale bars, 150 μm (**b**, **c**). **d**, Onset of neural crest migration; *x*-axis labels indicate migratory status of donor→host. **e**, Percentage of migrating neural crests at 10 h post-graft. Histograms show mean, error bars represent s.d. (**d**) or s.e.m. (**e**); one-way analysis of variance (ANOVA), $P < 0.0001$; two-tailed *t*-test, $^{***}P < 0.0002$, $^{****}P < 0.0001$, CI = 95%, $n = 43$ embryos in **b**–**e**). **f**, **i**, Time colour-coded trajectories and quantification of speed of neural crest migration towards Sdf-1. **f**, Non-migratory versus pre-migratory neural crest. **i**, Pre-migratory neural crest plated on soft or stiff substrates. Box plots in **f** and **i** show the median, box edges represent the 25th and 75th percentiles, and whiskers show spread of data (two-tailed *t*-test, $^{****}P < 0.0001$; $n = 78$ cells (**f**); $n = 79$ cells (**i**)). **g**–**h**, *In vivo* atomic force microscopy (iAFM) measurements. **g**, iAFM measurement direct on mesoderm. **i**, Spread of data for each stage, green lines show median, red whiskers represent interquartile range (two-tailed Mann–Whitney *U*-test, $^{****}P < 0.0001$, CI = 95%; $n = 259$ (stage 13), $n = 236$ (stage 17) and $n = 461$ (stage 20) AFM indentations; n = number of embryos; δ = mean indentation depth). Scale bars, 100 μm (**f**, **i**). **b**, **c**, **f** and **i** show representative examples from three independent experiments; CI = 95%.

lead to strain-stiffening of the tissue, increasing its apparent elastic modulus[12,17–19]. To reduce the stiffness experienced by neural crest cells, we mechanically ablated the surrounding tissue, leading to a release of tension and hence to a decrease in strain stiffening. Tissue ablation in the anterior region of the embryo (Fig. 2a) led to reduced mesodermal stiffness compared to the non-ablated side, as recorded by iAFM (Fig. 2b), and inhibited neural crest migration (Fig. 2c, d; ablation controls in Extended Data Fig. 4a, b).

Next, we carried out molecular manipulation of mesoderm stiffness by targeted injections of inhibitors of myosin activity, which decreased tissue stiffness (Fig. 2e, f and Extended Data Fig. 4c, d). The injection of either a translation-blocking morpholino against myosin light chain 9 (*myl9*-MO) or a constitutively active form of myosin phosphatase (Mypt1(T696A), referred here as CA-MYPT) markedly decreased mesoderm stiffness and blocked neural crest migration, as shown by *in situ* hybridization of the migratory neural crest marker *snail2* (also known as *snai2*) (Fig. 2f–i). Similar inhibition of neural crest migration was observed in embryos in which wild-type neural crests were grafted into embryos injected with *myl9*-MO or CA-MYPT (Extended Data Fig. 5a–d; fibronectin deposition control in Extended Data Fig. 6a–g), ruling out an effect of the manipulation of myosin on the migratory capacities of the neural crest cells. These data show that a critical stiffness in the mesoderm is required to promote neural crest migration *in vivo*.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

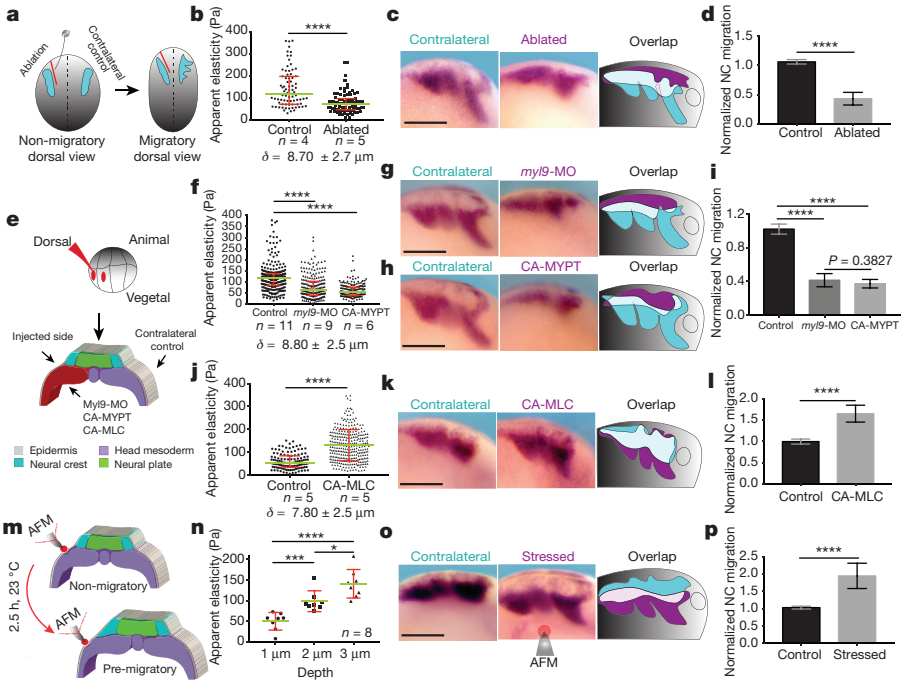


Figure 2 | Mesodermal stiffening is essential for neural crest CCM *in vivo*. **a**–**d**, Ablation experiments. **a**, Schematic showing ablation at non-migratory stages, and neural crest migration at stage 23 (migratory). **b**, iAFM measurements. $n = 78$ (control) and $n = 86$ (ablated) AFM indentations. **c**, Lateral views of embryos hybridized with a probe against *snail2* to analyse neural crest migration. **d**, Normalized neural crest migration ($n = 10$ embryos). **e**–**l**, Mesoderm-targeted injections. **e**, Embryos were injected into two dorso-vegetal blastomeres (prospective mesoderm). **f**, **j**, iAFM measurements. In **f**, $n = 294$ (control), $n = 224$ (*myl9*-MO), $n = 223$ (CA-MYPT); in **j**, $n = 120$ (control), $n = 301$ (CA-MLC). **g**, **h**, **k**, *snail2*-hybridized embryos. **i**, **l**, Normalized neural crest migration. $n = 25$ embryos (**i**); $n = 12$ embryos (**l**). **m**–**p**, Compression experiments. **m**, Schematic of AFM compression. **n**, Apparent elasticity plotted as a function of indentation depth ($n = 8$ embryos), green lines show median and whiskers represent the spread of data (excluding outliers). **o**, *snail2*-hybridized embryos. **p**, Normalized neural crest migration ($n = 13$ embryos). Overlap drawings in **c**, **g**, **h**, **k** and **o** are shown to facilitate comparison of neural crest migration, control neural crest (cyan) and treated neural crest (magenta). **b**, **f**, **j**, Direct iAFM measurements on mesoderm; green lines show median, red whiskers represent interquartile range; two-tailed Mann–Whitney *U* test, $****P < 0.0001$, CI = 95%, n = number of embryos, δ = mean maximum indentation depth. Histograms in **d**, **i**, **l** and **p** show mean, error bars represent s.d.; one-way ANOVA, $P < 0.0001$; two-tailed *t*-test, $*P = 0.014$, $***P < 0.001$, $****P < 0.0001$, CI = 95%). **c**, **g**, **h**, **k** and **o** show representative examples from three independent experiments. Scale bars, 200 μm.

We next investigated whether reaching this threshold stiffness in non-migratory embryos was sufficient to trigger premature neural crest migration. Injection of a constitutively active form of myosin light chain (DD-MLC, referred here as CA-MLC) into the mesoderm led to an early increase in tissue stiffness equivalent to that reached in pre-migratory embryos, and promoted premature neural crest migration (Fig. 2j–l; Extended Data Fig. 5e–g). To increase tissue stiffness without genetic or pharmacological manipulations, we applied locally a sustained compressive force on the embryos using iAFM (Fig. 2m; compression controls in Extended Data Fig. 4e–i), leading to strain stiffening[12,18,19] and premature neural crest migration (Fig. 2n–p). These results demonstrate that mesodermal stiffening is necessary and sufficient to trigger neural crest CCM *in vivo*.

To investigate how neural crest cells sense this increase in mesodermal stiffness, we perturbed signalling by inhibiting the integrin–vinculin–talin complex, which mediates mechanosensing in other systems[20]. We inhibited the integrin pathway in the neural crest by using a morpholino against *itgb1* (*itgb1*-MO) and dominant-negative forms of vinculin and talin. Inhibition of any of these mechanosensors led to a strong impairment of the onset of neural crest migration *in vivo* (Extended Data Fig. 7a–c), without significantly affecting attachment *in vitro*. These results show that the neural crest requires integrin–vinculin–talin for its migration, suggesting that this pathway is involved in mechanosensing by neural crest cells.

Our next goal was to examine the mechanisms underlying mesodermal stiffening *in vivo*. Modifications in ECM (for example, an increase in collagen fibres) have been proposed to affect tissue stiffness[21,22]; however, collagen is not expressed in the tissue during neural crest migration[23]. Together with our data that rules out fibronectin as a source of stiffening (Extended Data Fig. 2a–f), this suggests that changes in ECM are unlikely to be the cause of mesoderm stiffening in this system.

Although it has been proposed that myosin contractility increases tension locally, and thus stiffness of the dorsal mesoderm[13], our control experiments (Extended Data Figs 8a–e; 9a–h) suggest that actomyosin contractility is not essential for controlling head mesodermal stiffness during the stages at which the neural crest migrates. However, when we inhibited myosin in non-migratory embryos, mesoderm stiffness was reduced and neural crest migration was impaired (Fig. 2f–i and Extended Data Fig. 9a–c). This temporal difference in action is likely to be because myosin activity is required for the migration of mesoderm cells between non- and pre-migratory stages[13,24], which in turn indirectly affects subsequent mesoderm stiffness.

Another factor that controls tissue stiffness is cell density[12,25,26]. During development, mesoderm cells accumulate and intercalate in dorsal regions via convergent extension[24]. To test whether mesodermal cells accumulate underneath the neural crest, we measured mesodermal cell densities from non-migratory to pre-migratory stages. Cell densities increased towards the onset of neural crest migration (Fig. 3a–c), and furthermore, cell density directly correlated with mesoderm stiffening in all the treatments that affected migration ($R = 0.86$, $n = 16$ embryos) (Fig. 3d). These observations confirm that once cells have accumulated by convergent extension, the main contributor to mesoderm stiffness is cell density.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

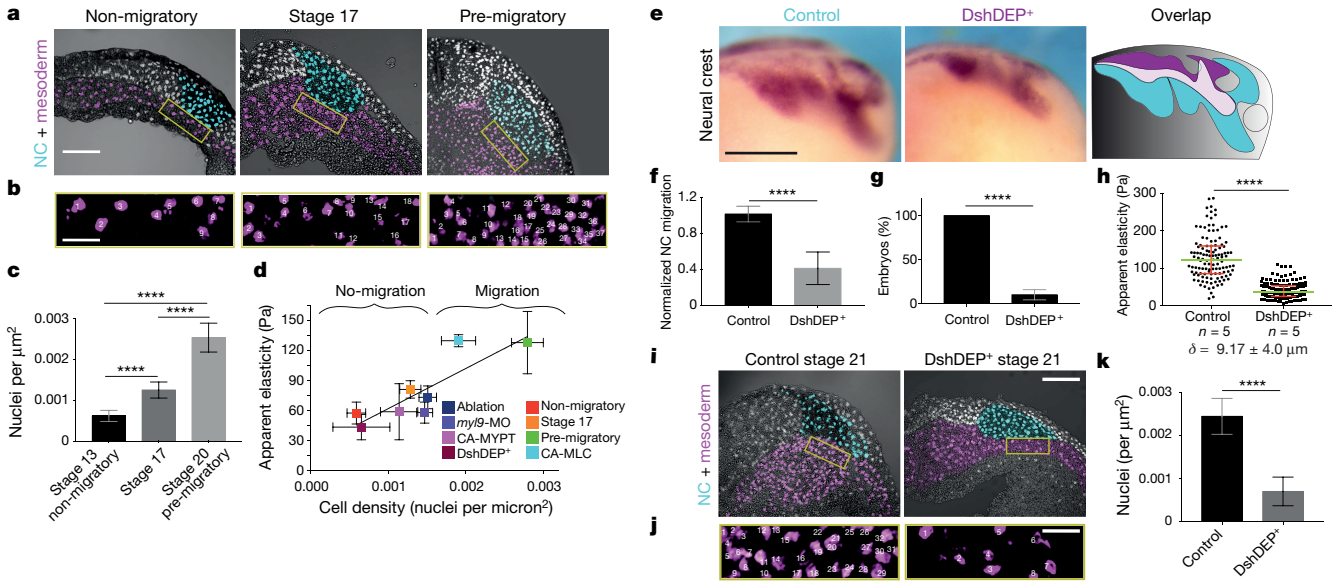


Figure 3 | PCP-dependent convergent extension is the driving force for mesodermal stiffening. a–d, Cell density analysis. **a**, Cross-sections of embryos at the indicated stages showing embryo morphology and nuclei pseudo-stained for neural crest (cyan) and mesoderm (magenta). **b**, Magnification of the areas outlined in **a**, used to quantify cell density. **c**, Nucleus density under the neural crest ($n = 27$ sections). **d**, Apparent elasticity of the head mesoderm as a function of cell density (Pearson test, $R = 0.8627$, $n = 16$ embryos). **e–k**, Effect of DshDEP$^+$. **e**, Lateral views and overlaps of control (cyan) or DshDEP$^+$ (magenta) *snail2*-hybridized embryos. **f**, Normalized neural crest migration ($n = 14$ embryos). **g**, Percentage of embryos displaying neural crest migration. **h**, Direct iAFM measurement on mesoderm. Green lines show median, red whiskers represent interquartile range (two-tailed Mann–Whitney *U*-test, $^{****}P < 0.0001$, CI = 95%; $n = 120$ (control), $n = 148$ embryos (DshDEP$^+$), δ = mean maximum indentation depth). **i**, Overlay showing embryo morphology and nuclei. **j**, Magnification of areas outlined in **i**, used to quantify cell density. **k**, Nucleus density under the neural crest, ($n = 11$ embryos). Histograms in **c**, **f**, **g** and **k** show mean; error bars, s.d. (**c**, **f**, **k**) or s.e.m. (**g**). One-way ANOVA, $P < 0.0001$; two-tailed *t*-test, $^{****}P < 0.0001$, CI = 95%. Scale bars: 100 µm (**a, i**), 50 µm (**b, j**) 200 µm (**e**). **a**, **b**, **e**, **i** and **j** show representative examples from three independent experiments.

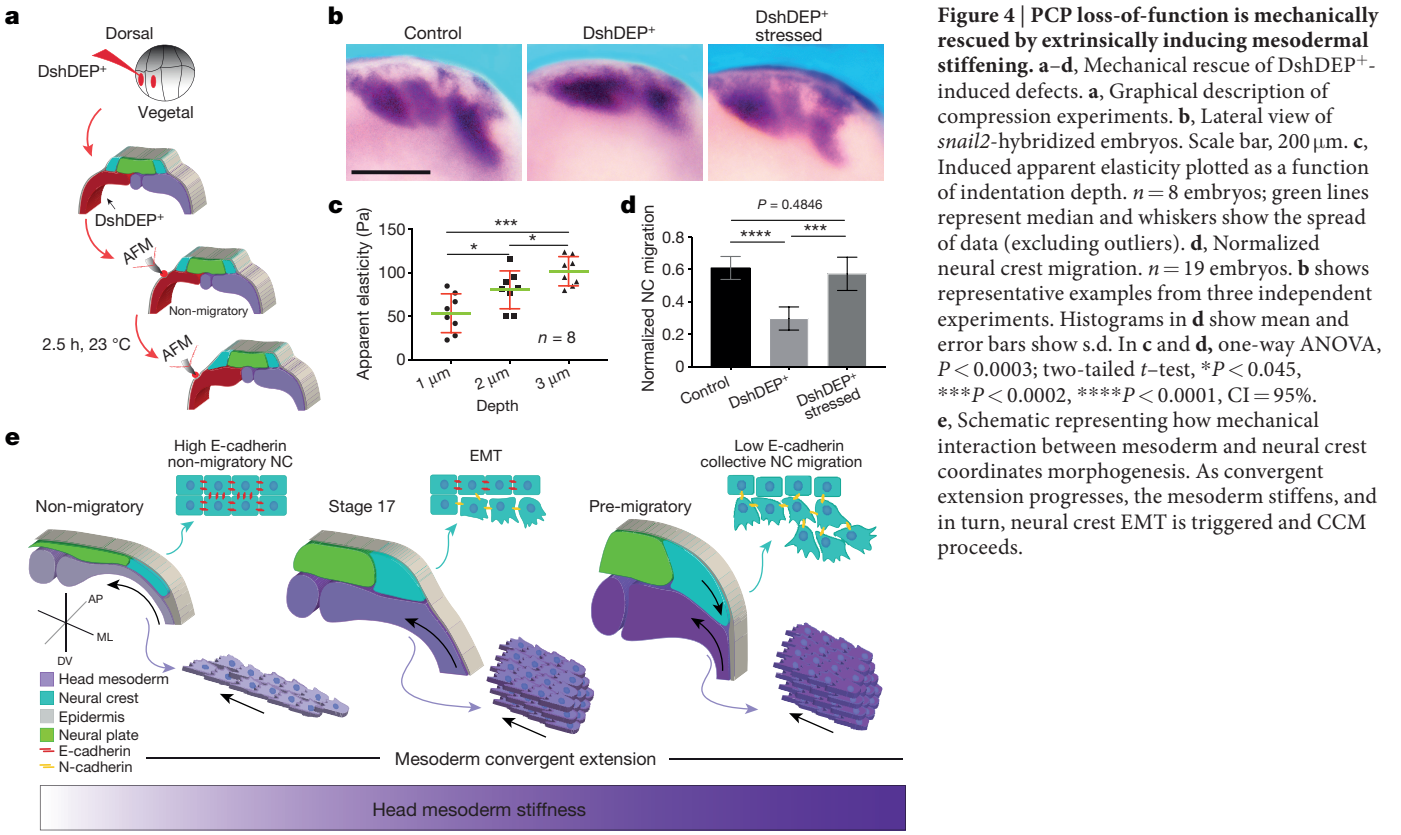


Figure 4 | PCP loss-of-function is mechanically rescued by extrinsically inducing mesodermal stiffening. a–d, Mechanical rescue of DshDEP$^+$-induced defects. **a**, Graphical description of compression experiments. **b**, Lateral view of *snail2*-hybridized embryos. Scale bar, 200 µm. **c**, Induced apparent elasticity plotted as a function of indentation depth. $n = 8$ embryos; green lines represent median and whiskers show the spread of data (excluding outliers). **d**, Normalized neural crest migration. $n = 19$ embryos. **b** shows representative examples from three independent experiments. Histograms in **d** show mean and error bars show s.d. In **c** and **d**, one-way ANOVA, $P < 0.0003$; two-tailed *t*–test, $^{*}P < 0.045$, $^{***}P < 0.0002$, $^{****}P < 0.0001$, CI = 95%. **e**, Schematic representing how mechanical interaction between mesoderm and neural crest coordinates morphogenesis. As convergent extension progresses, the mesoderm stiffens, and in turn, neural crest EMT is triggered and CCM proceeds.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

To validate the role of convergent extension in mesoderm stiffening, we specifically inhibited the planar cell polarity pathway (PCP) in the mesoderm with targeted injections of the PCP inhibitor DshDEP$^+$ (ref. 24). Inhibition of PCP signalling impaired neural crest migration in a non-autonomous manner, as shown by *in situ* hybridization and graft experiments (Fig. 3e–g and Extended Data Fig. 5h–k; fibronectin controls in Extended Data Fig. 6c, f). In agreement with our hypothesis, cell densities and mesoderm stiffness were reduced in DshDEP$^+$-injected embryos (Fig. 3h–k).

To confirm that the inhibition of neural crest migration after blocking PCP is a consequence of a mechanical change in the tissue, we rescued mesodermal stiffness in the softer PCP-depleted embryos by applying sustained extrinsic compression with iAFM[12,18,19] (Fig. 4a). This was sufficient to completely rescue neural crest migration (Fig. 4b–d), and to promote premature neural crest migration (Extended Data Fig. 10). These results indicate that the impaired neural crest migration observed in PCP-depleted embryos was mainly due to the decreased mesodermal stiffness.

While previous *in vitro* studies have proposed that mechanical cues may promote EMT[10] and CCM[11], we show here that convergent extension of the mesoderm leads to local stiffening, which in turn works as a long-range signal, triggering EMT and CCM of the neural crest *in vivo* (Fig. 4e). It is well known that tissue mechanics are a key determinant in many biological processes, such as axon growth[12], central nervous system development[25] and maintenance of tissue integrity via cell–cell adhesion during collective cell movement[15,16,27,28]. The most common regulators of tissue stiffness include ECM accumulation[21,22], cortical contractility[13,16,17] and cell densities[12,25,26]. Here we identified cellular rearrangements driven by PCP-dependent convergent extension as a mechanism that increases cell densities to increase stiffness in a developing organism. Similar changes in tissue stiffness are likely to occur in other systems undergoing PCP-dependent convergent extension.

An exciting new area of research is related to the *in vitro* generation of organ-analogous structures called organoids[4,29,30]. This research has focused mainly on the molecular manipulation of homogenous cell populations and mechanical inputs are just starting to be considered in the field[29,30]. As our findings reveal that proper morphogenesis requires mechanical interactions between different tissues, we anticipate that the consideration of mechanical signals in the organoid field will increase the success of *ex vivo*-engineered tissues. Moreover, as CCM is a common mode of migration during development, regeneration, wound healing and cancer progression[1,2,15,16,24,28], it is tempting to speculate that mechanical cues from the surrounding tissues could also play an important role during CCM in these systems.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 March 2017; accepted 11 January 2018.
Published online 14 February 2018.

1. Roca-Cusachs, P., Sunyer, R. & Trepat, X. Mechanical guidance of cell migration: lessons from chemotaxis. *Curr. Opin. Cell Biol.* **25,** 543–549 (2013).
2. Mayor, R. & Etienne-Manneville, S. The front and rear of collective cell migration. *Nat. Rev. Mol. Cell Biol.* **17,** 97–109 (2016).
3. Kerosuo, L. & Bronner-Fraser, M. What is bad in cancer is good in the embryo: importance of EMT in neural crest development. *Semin. Cell Dev. Biol.* **23,** 320–332 (2012).
4. Gilmour, D., Rembold, M. & Leptin, M. From morphogen to morphogenesis and back. *Nature* **541,** 311–320 (2017).
5. Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **166,** 21–45 (2016).
6. Betancur, P., Bronner-Fraser, M. & Sauka-Spengler, T. Assembling neural crest regulatory circuits into a gene regulatory network. *Annu. Rev. Cell Dev. Biol.* **26,** 581–603 (2010).
7. Nieuwkoop, P. D. & Faber, J. *Normal Table of* Xenopus laevis *(Daudin): a Systematical and Chronological Survey of the Development from the Fertilized Egg till the End of Metamorphosis* 2nd edn (North-Holland, 1967).
8. Theveneau, E. *et al.* Collective chemotaxis requires contact-dependent cell polarity. *Dev. Cell* **19,** 39–53 (2010).
9. Alfandari, D., Cousin, H., Gaultier, A., Hoffstrom, B. G. & DeSimone, D. W. Integrin $\alpha5\beta1$ supports the migration of *Xenopus* cranial neural crest on fibronectin. *Dev. Biol.* **260,** 449–464 (2003).
10. Wei, S. C. *et al.* Matrix stiffness drives epithelial–mesenchymal transition and tumour metastasis through a TWIST1–G3BP2 mechanotransduction pathway. *Nat. Cell Biol.* **17,** 678–688 (2015).
11. Sunyer, R. *et al.* Collective cell durotaxis emerges from long-range intercellular force transmission. *Science* **353,** 1157–1161 (2016).
12. Koser, D. E. *et al.* Mechanosensing is critical for axon growth in the developing brain. *Nat. Neurosci.* **19,** 1592–1598 (2016).
13. Zhou, J., Kim, H. Y. & Davidson, L. A. Actomyosin stiffens the vertebrate embryo during crucial stages of elongation and neural tube closure. *Development* **136,** 677–688 (2009).
14. Barriga, E. H., Maxwell, P. H., Reyes, A. E. & Mayor, R. The hypoxia factor Hif-1α controls neural crest chemotaxis and epithelial to mesenchymal transition. *J. Cell Biol.* **201,** 759–776 (2013).
15. Rauzi, M., Verant, P., Lecuit, T. & Lenne, P. F. Nature and anisotropy of cortical forces orienting *Drosophila* tissue morphogenesis. *Nat. Cell Biol.* **10,** 1401–1410 (2008).
16. Bjerke, M. A., Dzamba, B. J., Wang, C. & DeSimone, D. W. FAK is required for tension-dependent organization of collective cell movements in *Xenopus* mesendoderm. *Dev. Biol.* **394,** 340–356 (2014).
17. Wang, N. *et al.* Mechanical behavior in living cells consistent with the tensegrity model. *Proc. Natl Acad. Sci. USA* **98,** 7765–7770 (2001).
18. Storm, C., Pastore, J. J., MacKintosh, F. C., Lubensky, T. C. & Janmey, P. A. Nonlinear elasticity in biological gels. *Nature* **435,** 191–194 (2005).
19. Pogoda, K. *et al.* Compression stiffening of brain and its effect on mechanosensing by glioma cells. *New J. Phys.* **16,** 075002 (2014).
20. Charras, G. & Sahai, E. Physical influences of the extracellular environment on cell migration. *Nat. Rev. Mol. Cell Biol.* **15,** 813–824 (2014).
21. Moeendarbary, E. *et al.* The soft mechanical signature of glial scars in the central nervous system. *Nat. Commun.* **8,** 14787 (2017).
22. Chevalier, N. R. *et al.* How tissue mechanical properties affect enteric neural crest cell migration. *Sci. Rep.* **6,** 20927 (2016).
23. Seufert, D. W., Hanken, J. & Klymkowsky, M. W. Type II collagen distribution during cranial development in *Xenopus laevis*. *Anat. Embryol. (Berl.)* **189,** 81–89 (1994).
24. Shindo, A. & Wallingford, J. B. PCP and septins compartmentalize cortical actomyosin to direct collective cell movement. *Science* **343,** 649–652 (2014).
25. Koser, D. E., Moeendarbary, E., Hanne, J., Kuerten, S. & Franze, K. CNS cell distribution and axon orientation determine local spinal cord mechanical properties. *Biophys. J.* **108,** 2137–2147 (2015).
26. Shyer, A. E. *et al.* Emergent cellular self-organization and mechanosensation initiate follicle pattern in the avian skin. *Science* **357,** 811–815 (2017).
27. Dzamba, B. J., Jakab, K. R., Marsden, M., Schwartz, M. A. & DeSimone, D. W. Cadherin adhesion, tissue tension, and noncanonical Wnt signaling regulate fibronectin matrix organization. *Dev. Cell* **16,** 421–432 (2009).
28. Weber, G. F., Bjerke, M. A. & DeSimone, D. W. A mechanoresponsive cadherin–keratin complex directs polarized protrusive behavior and collective cell migration. *Dev. Cell* **22,** 104–115 (2012).
29. Wrighton, P. J. & Kiessling, L. L. Forces of change: mechanics underlying formation of functional 3D organ buds. *Cell Stem Cell* **16,** 453–454 (2015).
30. Poh, Y. C. *et al.* Generation of organized germ layers from a single mouse embryonic stem cell. *Nat. Commun.* **5,** 4000 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. J. Thompson for assistance with the AFM. This study was supported by grants to R.M. from the Medical Research Council (J000655), Biotechnology and Biological Sciences Research Council (M008517) and Wellcome Trust; to K.F. from the Medical Research Council (Career Development Award G1100312/1) and from the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (R21HD080585); to G.C. from the European Research Council (Consolidator grant MolCellTissMech, agreement 647186); and by postdoctoral fellowships to E.H.B. from EMBO (LTF-971) and Marie Skłodowska Curie (IF-2014_ST VivoMechCollMigra, agreement 658536). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions E.H.B. and R.M. conceived the project, E.H.B., K.F., G.C. and R.M. designed the experiments. All the experiments and analyses were performed by E.H.B. E.H.B. and R.M. wrote the manuscript. All the authors edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to R.M. (r.mayor@ucl.ac.uk).

Reviewer Information *Nature* thanks L. A. Davidson, N. Wang and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

***Xenopus* manipulation, *in situ* hybridization and *in vitro* transcription.** *Xenopus* embryos were obtained as previously described[14]. In brief, ovulation of mature females was induced by injecting serum gonadotrophin (Intervet) and chorionic gonadotrophin (Intervet). Released eggs were collected and fertilized *in vitro* by mixing with a sperm solution. Embryonic stages were determined as described[7]. For *in situ* hybridizations, an antisense digoxigenin-labelled RNA probe against *snail2*[31], a neural crest marker, was used and embryos at non-migratory, pre-migratory or migratory stages were fixed, hybridized and stained using established protocols[32]. Templates for riboprobes and mRNAs were generated as previously described[14,32]. In brief, for mRNA transcription an mMESSAGE mMACHINE SP6 Transcription Kit (Thermo-Fisher AM1340) was used. The riboprobe against *snail2* was transcribed with a Riboprobe *in vitro* Transcription System (Promega P1420). All animal experiments were approved by the Biological Service Unit at University College London and complied with UK Home Office guidelines (Animal Act 1986).

Morpholino and mRNA injections and reagents. Microinjections were performed using calibrated needles as described[33]. Fertilized eggs were de-jellied for 5 min with a solution containing 1 g cysteine (Sigma) and 1 ml of 5N NaOH in 50 ml ddH_2O. For cell labelling, 10 nl nuclear RFP (nRFP) and/or membrane GFP (mGFP) mRNAs were injected at the two-cell stage into one blastomere (250 pg per embryo per construct). Vinculin-Cter and talin-Nter were previously validated[34] and injected at the eight-cell stage in one dorsal blastomere (500 pg per embryo per construct). *itgb1*-MO (5′-GAATACTGGATAA- CGGGCCATCTTA-3′) was synthesized by GeneTools (GeneTools) and injected as previously described[35]. For targeted injections to the mesoderm, two dorso-vegetal blastomeres were injected (as shown in Fig. 2e and Extended Data Fig. 4c). Embryos were injected with 5 pg per blastomere of a morpholino against *myl9*[24] (*myl9*-MO 5′-TGGCTTTTGTCTTCTTGCTGGACAT-3′), synthesized by GeneTools (GeneTools); 1 ng per blastomere of $DshDEP^+$ (ref. 24); 1 ng per blastomere of CA-MYPT[36]; or 1 ng per blastomere of CA-MLC[37], both kindly provided by M. Tada.

For blebbistatin incubations, embryos were incubated as indicated in Extended Data Fig. 9, in a solution containing 100 μM blebbistatin (ab12042) dissolved in dimethyl sulfoxide (DMSO, Thermo-Fisher).

***Ex vivo* neural crest culture, chemotaxis assay on glass and graft experiments.** Neural crests were dissected at non-migratory or pre-migratory stages as previously described[14]. In brief, 20 min prior to dissection the chorion was removed with fine forceps, embryos were placed in a dish containing modelling clay, and a hair knife was used to remove the epidermis and explant the neural crest. After dissection, neural crests were maintained in Danilchik's medium (DFA; 53 mM NaCl, 5 mM Na_2CO_3, 4.5 mM KGluconate, 32 mM NaGluconate, 1 mM $MgSO_4(7H_2O)$, 1 mM $CaCl_2$, 0.1% BSA; adjusted to pH 8.3 with 1 M Bicine). To test the response of neural crests to Sdf-1 we used a chemotaxis assay[8,38] in which acrylic heparin beads (Sigma-Aldrich) were coated overnight with 1 μg/ml Sdf-1 (Sigma-Aldrich), washed with PBS and fixed to a fibronectin-coated glass-bottom dish (μ-Dish, 35 mm high, Ibidi) using silicon grease. Fluorescently labelled explants (mRFP or mGFP) were placed two diameters away from the bead and migration was recorded by time-lapse microscopy.

For graft experiments, wild-type or treated neural crests from donor and host embryos were removed as described above; a hair knife was used to locate the donor neural crest in the host embryo and a piece of glass coverslip was used to hold the grafted neural crest. The coverslip was removed after 1 h and the embryos were mounted for live imaging. Embryos were mounted on agarose dishes containing 1.2-mm lanes with 4% methyl cellulose (Sigma-Aldrich) and filled with DFA medium.

Polyacrylamide hydrogels: preparation and functionalization. Polyacrylamide (PAA) gels were prepared as follows. To allow adhesion to PAA, 76 mm × 24 mm glass slides were coated overnight at room temperature in a solution containing 14:1:1 ethanol:acetic acid:PlusONE Bind-Silene (GE Healthcare; vol:vol:vol). The slides were washed twice with 70% ethanol, air dried and stored at room temperature. For hydrophobic coating, 13-mm diameter × 0.1 mm glass coverslips were air cleaned and coated for 15 min at room temperature with PlusONE Repel-Silene ES (GE Healthcare), carefully recovered, air-dried and used immediately.

PAA mixes were prepared as follows. For soft gels, a mix containing 550 μl 7.6 mM HCl, 369.5 μl ddH_2O, 0.5 μl N,N,N′,N′-tetramethylethylenediamine (TEMED, Sigma), 16 μl bis-acrylamide (BioRad) and 60 μl acrylamide (BioRad) was prepared. For stiff gels, a mix containing 550 μl 7.6 mM HCl, 349.5 μl ddH_2O, 0.5 μl TEMED, 20 μl bis-acrylamide and 75 μl acrylamide was prepared. Polymerization was started by adding 5 μl of 10% ammonium persulfate (GE Healthcare) to each mix.

Soft or stiff gels were obtained by placing a 12-μl drop of PAA mix on an adhesive slide and covering with a hydrophobic coverslip. Polymerization was allowed to proceed for 25 min at room temperarture in a humidifier and the coverslip was carefully removed. Gels were washed three times for 10 min with 10 mM HEPES buffer.

Fibronectin was covalently linked to soft or stiff gels via Sulfo-SANPAH-mediated succinimide cross-linking. To activate gels, aliquots containing 1 mg of Sulfo-SANPAH (Thermo-Fisher 22589) in 20 μl anhydrous DMSO (Thermo-Fisher D12345) were rapidly diluted in 480 μl of 10 mM HEPES buffer. From the resulting solution, 25 μl was added to the gel surface and gels were irradiated with ultraviolet light for 5 min. The excess of Sulfo-SANPAH was removed by washing for 5 min with 10 mM HEPES buffer. A drop containing 10 mg/ml fibronectin was added to the surface of the activated gels, covered with Parafilm, and incubated for 2 h in a humidifier. The parafilm was removed and the functionalized gels were washed twice for 5 min in 10 mM HEPES buffer. Young's modulus (in Pa) for each gel was determined by AFM and functionalization was checked by immunostaining against fibronectin (Extended Data Fig. 2k–m).

Chemotaxis and dispersion assays on hydrogels with varying stiffness values. *Chemotaxis.* To test how substrate stiffness contributes to the response of neural crest cells towards Sdf-1, we developed a chemotaxis assay on PAA gels of varying stiffness. In brief, once coated with Sdf-1 as described for the glass chemotaxis assay, acrylic heparin beads (Sigma-Aldrich) were split into two halves and each half was gently indented in the surface of fibronectin-coated PAA hydrogels (no differences in gel stiffness were observed upon bead indentation, controlled with AFM). Labelled neural crest explants were placed two to three diameters in front of the beads and migration was registered by time-lapse microscopy.

Dispersion assay. As a read-out of EMT we used a cell dispersion assay on PAA hydrogels of varying stiffness. In brief, labelled neural crests were explanted into fibronectin-coated hydrogels and allowed to disperse. Dispersion was recorded by time-lapse microscopy.

Immunostaining and phalloidin staining. For phospho-myosin (ab2480, Abcam) and fibronectin (mAb 4H2 anti-FN, kindly provided by D. DeSimone) immunostaining in flat-mounted embryos, whole embryos were fixed using 2% TCA (trichloroacetic acid) in 0.3% PBS-T (PBS, 0.3% Triton X-100) for 20 min. Dorsal halves of fixed embryos were dissected using a 15°, 5-mm depth microsurgical knife (MSP 7516). Dissected halves were washed with PBS and blocked for 4 h with 10% normal goat serum (NGS). Antibodies were diluted at 1:100 (anti-myosin) and 1:500 (anti-fibronectin) in 10% NGS and incubated overnight at 4 °C, followed by three washes with 0.3% PBS-T (PBS, 0.3% Triton X-100). Alexa-fluor (Thermo-Fisher) secondary antibodies were diluted 1:350 in 10% NGS and incubated overnight at 4 °C and excess antibody was removed by washing three times with 0.3% PBS-T. Nuclei were stained with DAPI in a mixture with the secondary antibodies (1:1,000). Finally, embryos were washed overnight in methanol and flat-mounted for imaging in clearing mix (two volumes benzyl alcohol and one volume benzyl benzoate; Sigma-Aldrich). For E-cadherin and N-cadherin detection, explants were fixed with 4% formaldehyde in 0.2% PBS-T (PBS, 0.2% Triton X-100) for 12 min and blocked with 10% NGS for 2 h. The primary antibody (anti-E-cadherin 5D3 or anti-N-cadherin MNCD2, Developmental Studies Hybridoma Bank) was diluted at 1:100 in 10% NGS and incubated overnight at 4 °C. Explants were washed three times with 0.2% PBS-T and incubated overnight at 4 °C with secondary antibody, diluted at 1:350 in 10% NGS. DAPI was diluted at 1:1,000 and mixed with the secondary antibodies. To detect fibronectin on gels, functionalized gels were washed twice with PBS and incubated for 1 h at room temperature with anti-human fibronectin (Sigma S3648) diluted 1:500 in PBS. Gels were washed three times with PBS and incubated for 1 h at room temperature with 1:350 Alexa-fluor anti-Rabbit 488 and 1:1,000 DAPI. Gels were washed three times with PBS and maintained in 1% formaldehyde in PBS. Fluorescence intensity was analysed using the measurement tool from ImageJ.

Phalloidin (Thermo Fisher A12379) staining was carried out as follows: embryos were fixed in a mix containing 4% paraformaldehyde and 0.25% glutaraldehyde for 10 min, washed three times in 0.3% PBS-T for 10 min and incubated in a solution containing 5U phalloidin per ml PBS. Excess phallodin was washed with 0.3% PBS-T and embryos were dehydrated in isopropanol (10 min) and cleared and mounted in 2:1 benzyl benzoate:benzyl alcohol (Sigma).

Microscopy and time-lapse live-imaging. *Time-lapse imaging.* Images for chemotaxis and dispersion assays (on gels and glass) were acquired every 5 min at 18 °C using an upright microscope (DMR XA2, Leica) equipped with a motorized stage (Prior Scientific) and a camera (Orca-5G Hamamatsu). Filter wheels, camera, stage and shutters were controlled with SimplePCI software (Hamamatsu). A 20× objective (HCX APO L 20×/0.50 W U-V-I FWD = 3.50 mm, Leica) was used.

Confocal time-lapse imaging. Images were acquired every 16 s at 18 °C using an upright confocal integrated microscope system (SP8vis, Leica) with a 63× objective (HCX APO L 63×/0.90 W U-V-I CS2, FWD = 2.20 mm, Leica).

Live imaging. For graft experiments, time-lapse images were acquired every 10 min with an upright microscope (Eclipse 80i, Nikon) fitted with a motorized stage

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

(Prior Scientific) and a camera (ORCA-05G, Hamamatsu Photonics). Images were acquired at 18 °C with a 4× objective (CFI Plan Fluor 4×/0.13, WFD = 17.1 mm). Camera, stage, shutter and filter wheels were controlled with SimplePCI software (Hamamatsu).

In situ hybridization imaging. All images were captured at room temperature in agarose dishes containing PBS, using a dissecting microscope (MZ FL III, Leica) equipped with a camera (DFL420, Leica) and imaging software (IM50, Leica). Magnification was 3.2×.

Immunofluorescence imaging. All images were acquired at room temperature. Images in Fig. 3a, b, j, k and Extended Data Figs 2b, f; 4d, g; 6; 9e, f were acquired in an inverted confocal microscope (FluoView FV1000, Olympus) equipped with a digital camera (DP73, Olympus) and a 20× objective (UPLSAPO 20×/0.75 FWD = 0.6 mm, Olympus). Camera, filter wheels, and shutters were controlled by FluoView Software (Olympus). Confocal images in Extended Data Figs 2k and 3 were acquired using an upright confocal integrated microscope system (SP8vis, Leica) with a 63× objective (HC PL APO 63×/1.40 Oil CS2, FWD = 0.14 mm, Leica).

***In vivo* atomic force microscopy measurements.** All AFM measurements were carried out as previously described[12,21,25,40]. In brief, a JPK Nanowizard Cellhesion 200 (JPK Instruments AG), fitted with a x–y-motorized stage and mounted on an inverted optical microscope (Axio ObserverA1, Zeiss) was used. Customized cantilevers were used. For mesoderm AFM measurements, tipless silicon cantilevers (Arrow-TL1, NanoWorld) were mounted on the AFM device and spring constants were calculated using the thermal noise method[39]. Cantilevers with spring constants between 0.01 and 0.03 N/m were selected. Monodisperse polystyrene beads (diameter: 37.28 ± 0.34 μm; microParticles) were glued to the selected cantilevers[12,21,25]. Embryos were mounted on a clay modelling dish and the epidermis was carefully dissected using a hair knife. Force–distance curves were acquired every 15 μm in the region immediately lateral to the neural crest as shown in Fig. 1g and Extended Data Fig. 2g. A region of interest was defined using a custom-written script that controls the x–y-motorized stage (maximum indentation force: 7 nN, approach speed: 5 μm/s, data rate: 1,000 Hz). Measurements on intact embryos were carried out in the same territory, but embryos were not dissected. In these experiments, monodisperse polystyrene beads (diameter: 89.3 ± 1.1 μm; microParticles GmbH) were glued to tipless silicon cantilevers with spring constants of ~0.1 N/m (Shocon-TL; AppNano), and force–distance curves were acquired with a maximum indentation force of 50 nN, approach speed: 5 μm/s, data rate: 1,000 Hz, one measurement per embryo. For extrinsic compression experiments, stress was applied in the same region where measurements were taken (Fig. 2m) with 89.3 ± 1.1 μm beads as cantilever probes. A sustained force of 80 nN (control embryos) and 50 nN (DshDEP$^+$-injected embryos) was applied to compress the tissue. Force was applied until embryos reached pre-migratory (control and DshDEP$^+$) or migratory stages (DshDEP$^+$). Uncompressed time-matched embryos and the contralateral side of the compressed embryos were used as controls. An image was acquired for each measurement using an adapted AxioZoom V.16 system (Zeiss) connected to an Andor Zyla 4.2 CMOS camera. After compression, embryos were rapidly fixed for *in situ* hybridization. Displacement maps were generated using the ImageJ iterative PIV basic plugin.

Data analysis and image treatment. *AFM data.* Force–distance curves were analysed as previously described, using a customized Matlab routine based on the Hertz model for a spherical indenter[12,21,25]

$$F = \frac{4}{3} K \sqrt{r}\,\delta^{3/2} = \frac{4}{3}\frac{E}{1-\nu^2}\sqrt{r}\,\delta^{3/2}$$

with applied force F, Young's modulus E, Poisson's ratio ν, indenter radius r, indentation depth δ, and apparent elastic modulus $K = E/(1-\nu^2)$, which is referred to as 'apparent elasticity' in the text. Force–distance curves from the measurements taken in the region of interest were analysed for defined indentation depths of 3 μm, which has been shown to measure the mechanical properties of developing tissues[12,21,25,40]. The mean maximum indentation depth is included in each figure. Measurements for each experiment were pooled and statistically analysed. Representative examples of force curves with highlighted indentation depths (δ) are provided (Extended Data Fig. 2i, j). Mean indentation depth (δ) for each AFM experiment is included below the corresponding charts.

***In vivo* analysis of neural crest migration.** For *in situ* hybridized embryos, the length of the treated neural crest was normalized against the length of the contralateral control side of the embryo. For grafted embryos, the length of the neural crest was normalized against the total dorsoventral length of the embryo at the position of the neural crest. Lengths were obtained using the measurement tool from ImageJ and further analysed as described in Statistical analysis.

***Ex vivo* analysis of neural crest migration.** Manual tracking and chemotaxis tool plugins from ImageJ were used to calculate speeds of migration. For dispersion assays, an ImageJ-based custom-made Delaunay-triangulation plugin[13] was used to calculate the distance between neighbour cells (available upon request). Data were further analysed as described in Statistical analysis.

Cell density measurements. Confocal images of transverse sections stained for DAPI were projected in ImageJ. A region of interest (ROI) was selected in the mesoderm under the neural crest in embryos at non-migratory stages; the number of nuclei within this region was quantified and nucleus density was calculated as the number of nuclei within the ROI divided by the area of the ROI. The same ROI size was used to calculate density at stage 17 and pre-migratory stages. Cell density was expressed in nuclei per μm^2.

Image treatment. *z*-stacks, maximum projections and time-lapse movies were created using ImageJ software. Adjustment of display map levels, re-sizing, addition of scale bars and pseudocolour were applied with ImageJ and/or Adobe Photoshop. In Fig. 2c, g, h, k, o, and 3a (pre-migratory *in situ* hybridization panel) and Fig. 4b, the background was pseudocoloured in Photoshop.

Cryosectioning. Fixed embryos were washed twice for 5 min with phosphate buffer (0.2 M $NaH_2PO_4{\cdot}H_2O$ and 0.2 M K_2HPO_4, pH 7.4), and incubated for 2 h at room temperature with a solution containing 15% sucrose (wt/vol; Sigma-Aldrich) in phosphate buffer and for 1 h at 42 °C in a gelatin solution containing 8% gelatin (Sigma-Aldrich) and 15% sucrose in phosphate buffer (wt/vol). Embryos were oriented in gelatin solution and gelatin blocks containing the embryos were snap frozen at −80 °C with pre-chilled isopentane. Samples were sectioned in 20-μm slices using a cryostat (CM-3050S, Leica) and collected in SuperFrozen Slides (VWR International). The slides were dried for at least 6 h and the gelatin was removed by washing twice with PBS for 15 min. MOWIOL (EMD Millipore) was used as mounting medium.

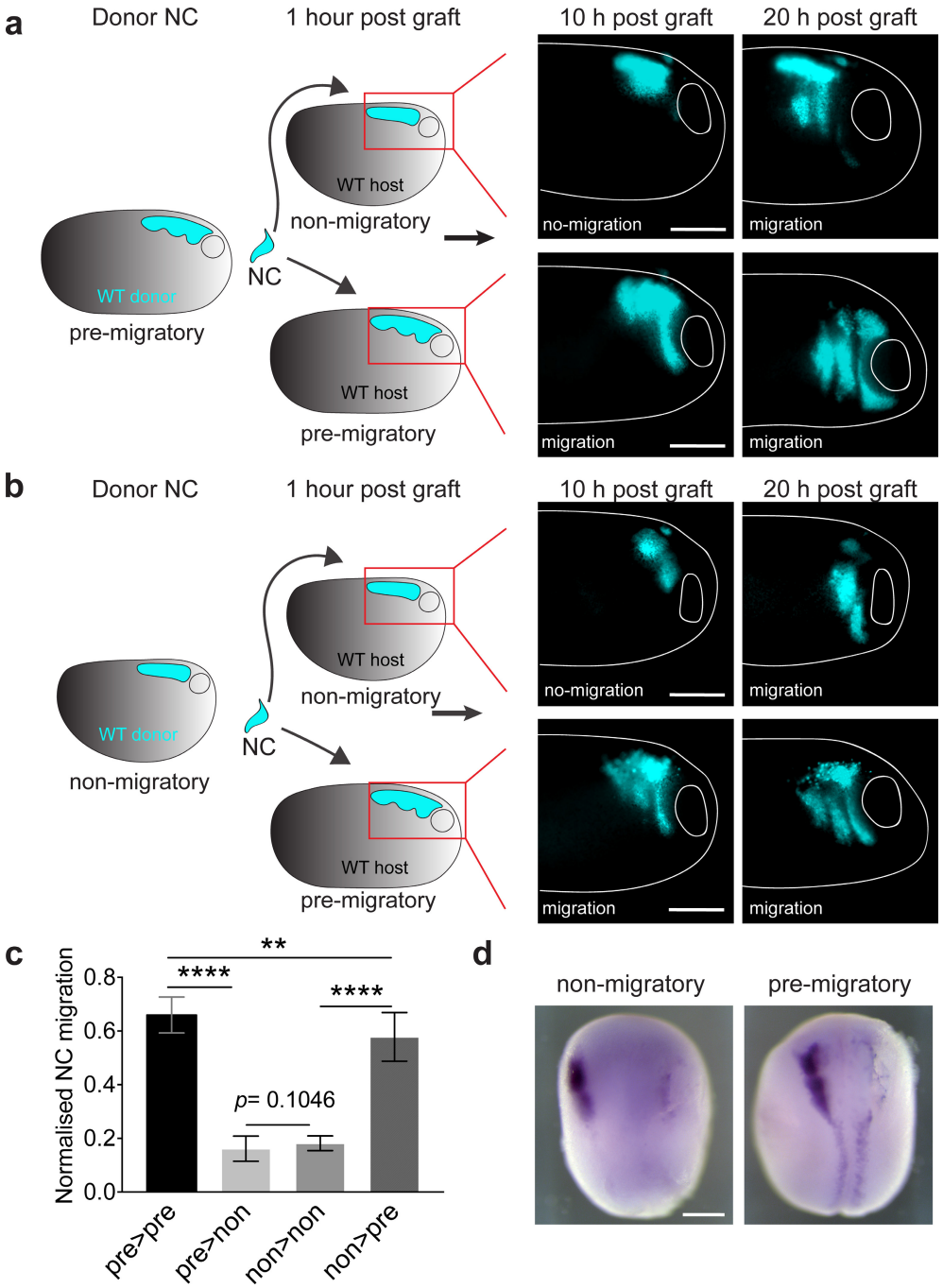
Statistical analysis. Sample size was determined by following published studies and no statistical method was used. The experiments were not randomized. Owing to the nature of the experiments, the authors were not blinded to allocation during experiments and results analysis, and only viable embryos and cell clusters were included for analysis. Additionally, mis-injected embryos were excluded from *in situ* hybridization analysis (correct injection was monitored by co-injection of linear tracing molecules mRFP or fluorescein dextran (FDX)). After selecting surviving and properly injected embryos and/or cells, our parameters for different experiments were measured at random.

Each experiment was repeated at least three times. Normality in the spread of data for each experiment was tested using the Kolmogorov–Smirnov, d'Agostino–Pearson or Shapiro–Wilk test in Prism7 (GraphPad). Significances for datasets displaying normal distributions were calculated in Prism7 with an unpaired Student's *t*-test (two-tailed, unequal variances) or one-way analysis of variance (ANOVA) for multiple comparisons. Individual comparisons were calculated only when ANOVA showed $P < 0.05$. Significances for non-normal distributed data were calculated in Prism7 using the Mann–Whitney *U*-test.

Data and code availability. Original data that support our findings and custom codes for analysis are available from the corresponding author upon reasonable request. Source Data for *P* values are provided in the online version of the paper.

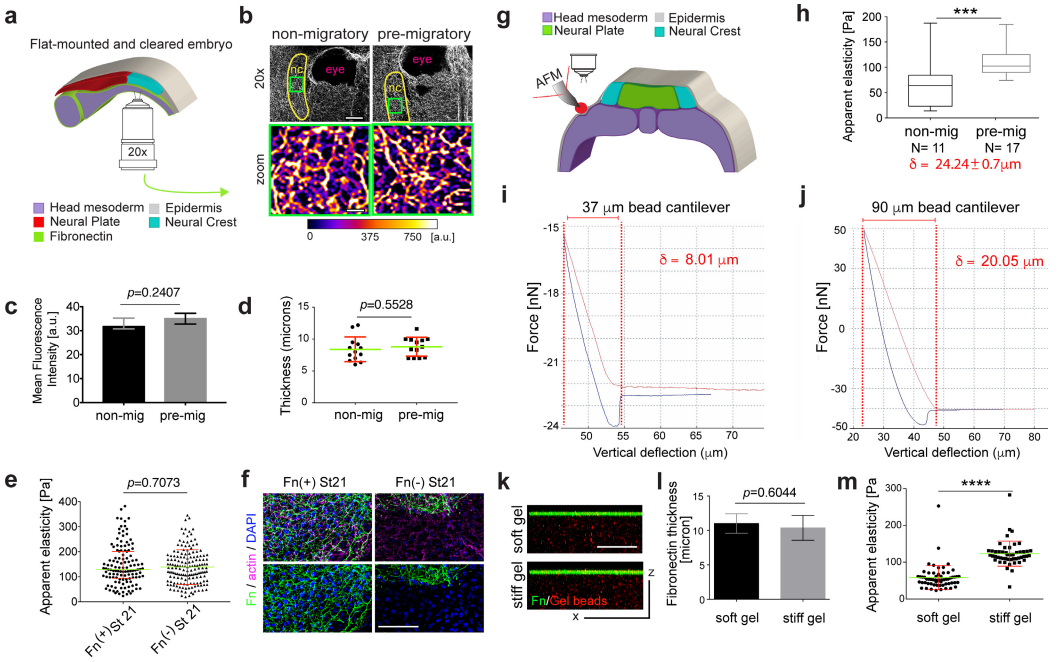
31. Mayor, R., Morgan, R. & Sargent, M. G. Induction of the prospective neural crest of *Xenopus*. *Development* **121,** 767–777 (1995).
32. Harland, R. M. *In situ* hybridization: an improved whole-mount method for *Xenopus* embryos. *Methods Cell Biol.* **36,** 685–695 (1991).
33. Mancilla, A. & Mayor, R. Neural crest formation in *Xenopus laevis*: mechanisms of *Xslug* induction. *Dev. Biol.* **177,** 580–589 (1996).
34. Iioka, H., Iemura, S., Natsume, T. & Kinoshita, N. Wnt signalling regulates paxillin ubiquitination essential for mesodermal cell motility. *Nat. Cell Biol.* **9,** 813–821 (2007).
35. Morita, H. *et al.* Cell movements of the deep layer of non-neural ectoderm underlie complete neural tube closure in *Xenopus*. *Development* **139,** 1417–1426 (2012).
36. Weiser, D. C., Row, R. H. & Kimelman, D. Rho-regulated myosin phosphatase establishes the level of protrusive activity required for cell movements during zebrafish gastrulation. *Development* **136,** 2375–2384 (2009).
37. Di Ciano-Oliveira, C. *et al.* Is myosin light-chain phosphorylation a regulatory signal for the osmotic activation of the Na^+–K^+–$2Cl^-$ cotransporter? *Am. J Physiol. Cell. Physiol.* **289,** C68 –C81 (2005).
38. Theveneau, E. & Mayor, R. Beads on the run: beads as alternative tools for chemotaxis assays. *Methods Mol. Biol.* **769,** 449–460 (2011).
39. Hutter, J. & Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **64,** 1868–1873 (1993).
40. Iwashita, M., Kataoka, N., Toida, K. & Kosodo, Y. Systematic profiling of spatiotemporal tissue and cellular stiffness in the developing brain. *Development* **141,** 3793–3798 (2014).
41. Benko, R. & Brodland, G. W. Measurement of in vivo stress resultants in neurulation-stage amphibian embryos. *Ann. Biomed. Eng.* **35,** 672–681 (2007).
42. Goto, T., Davidson, L., Asashima, M. & Keller, R. Planar cell polarity genes regulate polarized extracellular matrix deposition during frog gastrulation. *Curr. Biol.* **15,** 787–793 (2005).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



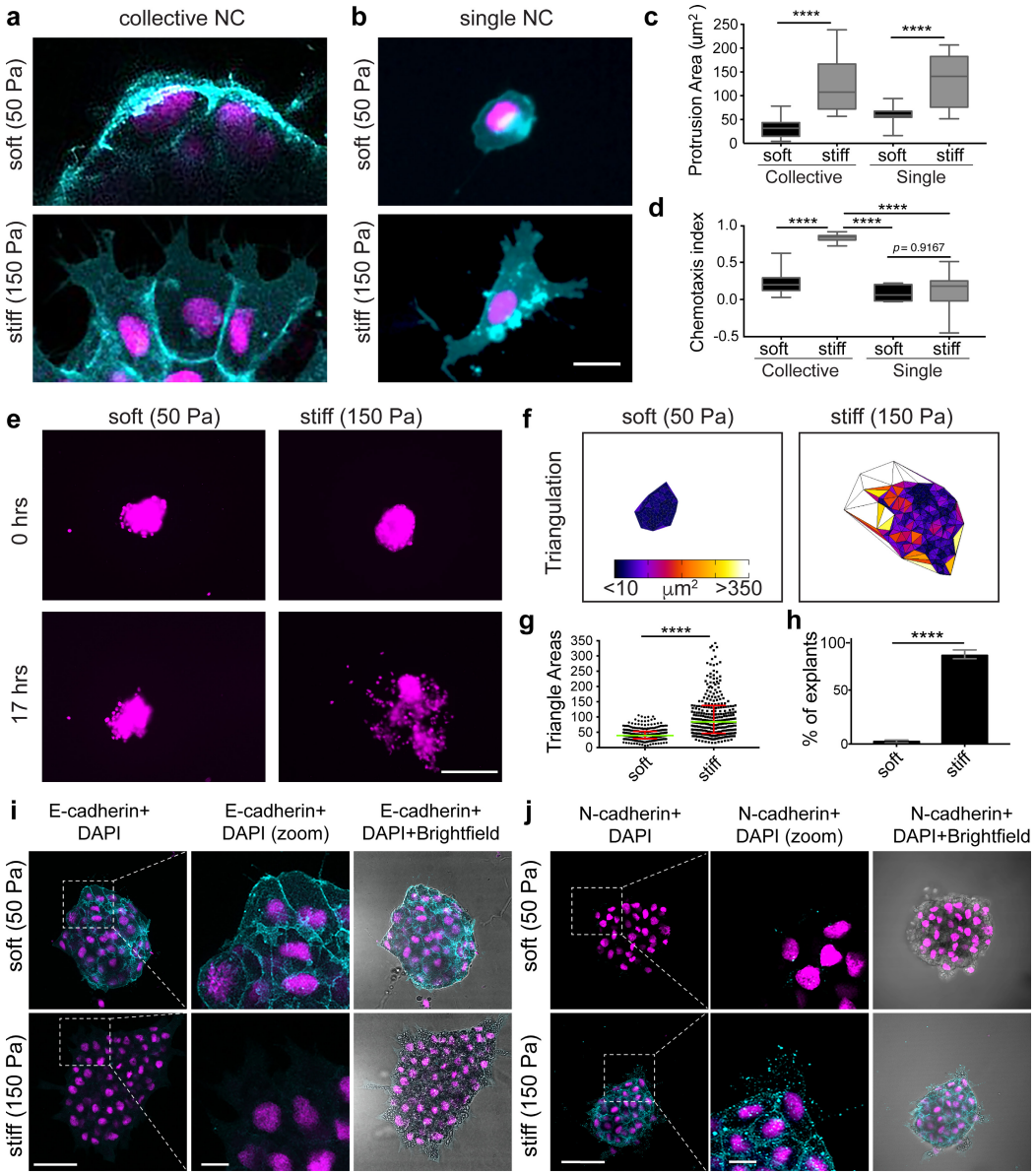
Extended Data Figure 1 | Neural crest migration is controlled by environmental factors. a–d, Heterochronic graft controls. In all grafts, we completely removed the host neural crest before implanting the graft, and all grafted cells migrated 20 h post graft, as expected for embryos reaching the migratory stage[7] (stage 23). **a**, Labelled pre-migratory neural crest grafted into unlabelled non- and pre-migratory hosts. Ten hours post graft, representative examples show that pre-migratory neural crest grafted into non-migratory hosts does not migrate and that pre-migratory neural crest grafted into pre-migratory hosts does migrate. **b**, Labelled non-migratory neural crest grafted into unlabelled non- and pre-migratory hosts. Ten hours post graft, representative examples show that non-migratory neural crests grafted into non-migratory hosts do not migrate and that pre-migratory neural crests grafted into pre-migratory hosts do migrate. Twenty hours post-graft panels in **a** and **b** show that neural crests migrated in all conditions after 20 h. **c**, Normalized neural crest migration. Histograms show mean, error bars represent s.d.; $n = 80$ measurements from 20 embryos (one-way ANOVA, $P < 0.0001$; two-tailed t-test, $^{**}P = 0.0026$, $^{****}P < 0.0001$, CI = 95%). **d**, Dorsal views of embryos hybridized with a probe against *snail2* after dissection of the host neural crest, showing that endogenous neural crest was completely removed before grafting. **a**, **b** and **d** show representative examples of three independent experiments; scale bar, 250 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



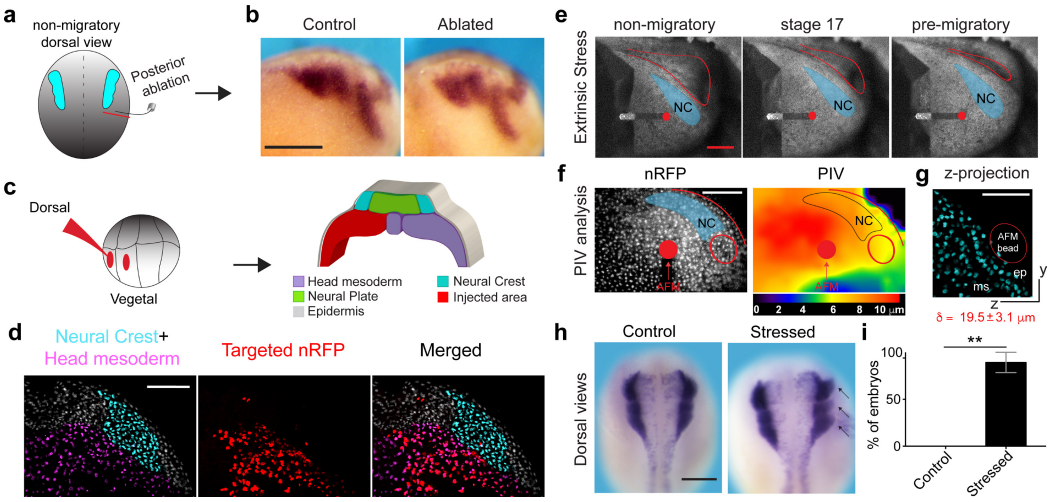
Extended Data Figure 2 | Fibronectin expression under the neural crest *in vivo* and contribution to tissue stiffness, stiffness measurements across epidermis, and *ex vivo* system characterization. a–d, Analysis of fibronectin expression under the neural crest. **a**, Schematic showing the imaging method used to access fibronectin under the neural crest. **b**, Immunostaining for fibronectin in embryos at non- and pre-migratory stages. 20× panels, yellow outline highlights the anatomical position of the neural crest; green box highlights the region shown in the zoom panels. Zoom panels show heat maps of fibronectin signal (representative of five independent samples). **c**, Fluorescence signal intensity. Histograms show mean, error bars represent s.d. (two-tailed *t*-test, CI = 95%). **d**, Fibronectin signal thickness in the mesoderm at pre- and non-migratory stages. Green lines show median and red whiskers represent interquartile range; $n = 10$ embryos (**c**), $n = 26$ sections from three independent experiments (**d**) (two-tailed *t*-test, CI = 95%). Scale bars: 50 μm (20× panels), 20 μm (zoom panels). **e**, **f**, Apparent elasticity was measured before and after mechanical removal of the ECM by gently scraping the surface of the tissue with a hair loop. **e**, Apparent elasticity was measured directly on the mesoderm in the presence ($Fn^{(+)}$) or absence ($Fn^{(-)}$) of fibronectin (Fn, green) in embryos at stage 21 (pre-migratory). Graphs plot spread of data; green lines show median; red whiskers represent interquartile range. Two-tailed Mann–Whitney *U*-test, *P* as indicated, CI = 95%; n = number of AFM indentations: $n = 124$ ($Fn^{(+)}$), $n = 157$ ($Fn^{(-)}$). Mean indentation depth = 8 μm. **f**, Representative immunostaining against fibronectin (Fn), conditions as indicated, $n = 7$ embryos. Scale bar, 100 μm. **g**, **h**, AFM controls. **g**, Schematic of iAFM measurement. To check that the dissection of the epidermis did not modify the mesodermal elasticity, we measured the elastic modulus through the epidermis, and no differences were found (compare with Fig. 1h). **h**, Apparent elasticity. Box plots show the median, box edges represent the 25th and 75th percentiles, and whiskers show the spread of data (excluding outliers). Two-tailed Mann–Whitney *U*-test, $^{***}P = 0.0006$, CI = 95%; n = number of embryos, δ = mean indentation depth. **i**, **j**, Representative examples of force curves obtained from measurements performed using cantilevers coated with either 37-μm beads (**i**, $n = 2386$) or 90-μm beads (**j**, $n = 28$). Red lines represent cantilever extension and dark blue lines represent retraction. Indentation depth (δ) for each curve is provided. **k**–**m**, *Ex vivo* system characterization. **k**, Cross-sectional confocal image of stiff or soft PAA hydrogels, confirming that gels (red) of varying stiffness are evenly functionalized with fibronectin (green). **l**, Fibronectin thickness does not change between soft and stiff gels. Histograms represent mean and bars show s.d.; $n = 8$ gels; two-tailed *t*-test, CI = 95%. Results in **k** and **l** are representative of gels from three independent experiments. **m**, AFM measurements. Spread of data, green lines show median, red whiskers represent interquartile range; two-tailed Mann–Whitney *U*-test, $^{****}P < 0.0001$, CI = 95%; $n = 60$ (stiff), $n = 50$ (soft) from 10 gels. Scale bars, 50 μm (**k**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



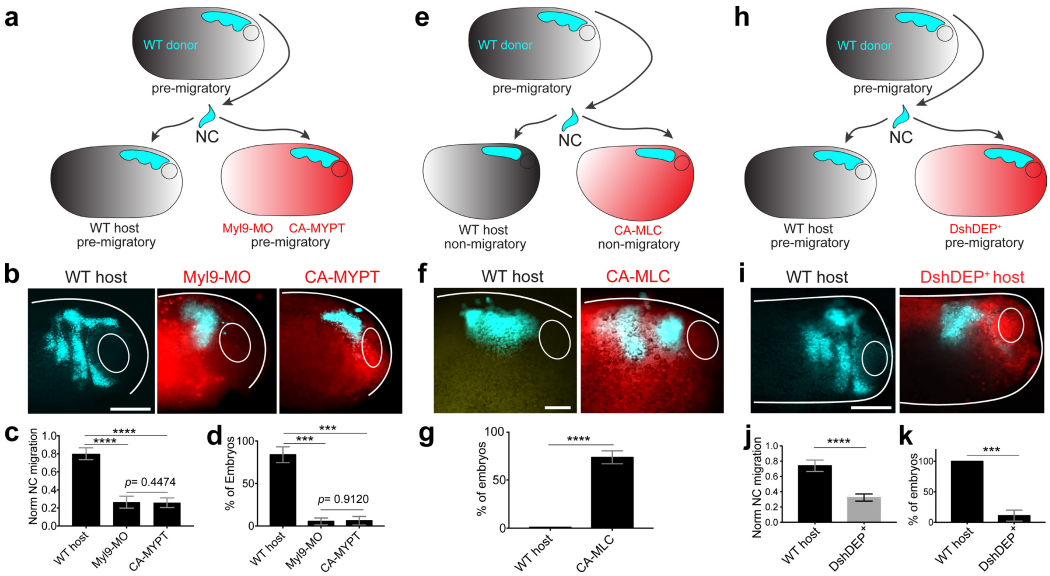
Extended Data Figure 3 | Neural crest motility, dispersion and EMT are controlled by substrate mechanics. a–d, Protrusion and chemotaxis analysis of collective and single neural crest cells plated on soft and stiff substrates. **a**, Confocal projections of labelled neural crest collectives plated on soft or stiff gels. **b**, Confocal projections of labelled single neural crest cells plated on soft or stiff gels. Protrusion area (**c**) and chemotaxis quantifications (**d**) for collective and single neural crest cells plated on soft or stiff gels, box plots show median, box edges represent 25th and 75th percentiles and whiskers show the spread of data (excluding outliers). $n = 63$ cells (protrusion area), $n = 79$ cells (chemotaxis); one-way ANOVA, $P < 0.0001$; two-tailed t-test, $^{****}P < 0.0001$, CI = 95%. **e–h**, Neural crest dispersion analysis on gels of varying stiffness. **e**, Neural crest cells labelled with nRFP and plated on stiff or soft gels are shown at 0 h and 17 h after plating. **f–h**, Quantification of cell dispersion. **f**, Colour-coded Delaunay triangulation shown at 17 h after plating to facilitate visualization of the distances between neighbour cells. **g**, Quantification of Delaunay triangulation. Green lines show median, red whiskers represent the interquartile range; two-tailed Mann–Whitney U-test, $^{****}P < 0.0001$, CI = 95%; $n = 730$ triangles from 31 explants. **h**, Percentage of dispersing explants. Histograms show mean and error bars represent s.e.m.; two-tailed t-test, $^{****}P < 0.0001$, CI = 95%; $n = 31$ explants. **g** and **h** show data from four independent experiments (soft) and three independent experiments (stiff). **i**, **j**, Immunostaining for E-cadherin (**i**) and N-cadherin (**j**) on neural crests plated on substrates of varying stiffness. **a**, **b**, **e**, **i** and **j** show representative examples from three independent experiments. Scale bars: 20 μm (**a**, **b**), 250 μm (**e**), 50 μm (**i**, **j**, main), 10 μm (**i**, **j**, zoom).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



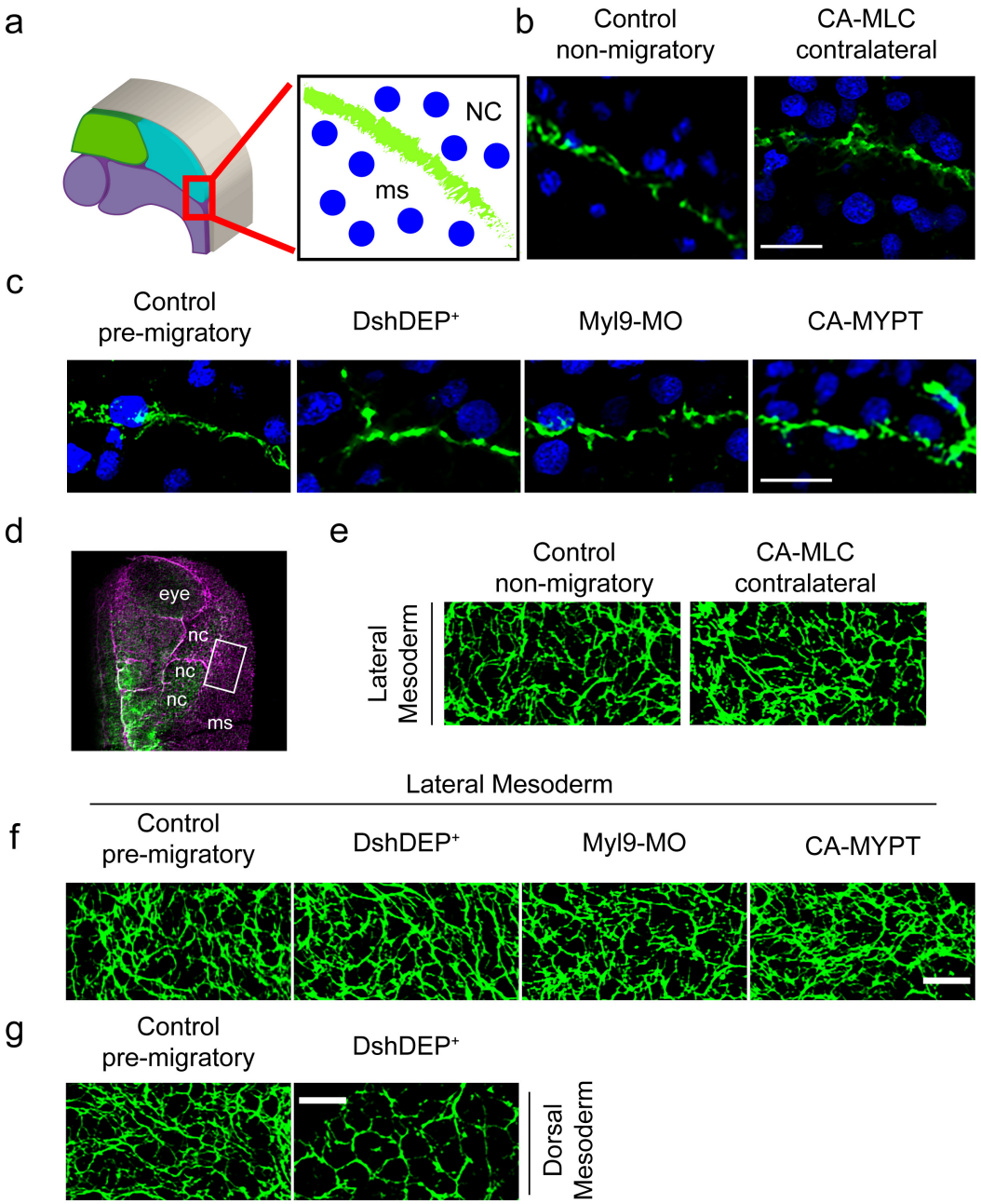
Extended Data Figure 4 | Posterior ablations, mesoderm targeted injections and extrinsic compression characterization. a, **b**, Ablation controls. Mechanical tension in *Xenopus* is higher in the anterior region of the embryo than in posterior territories[12,41]. Consistent with this observation, no effect on neural crest migration or tissue stiffness was observed when the ablation was done in posterior regions. **a**, Schematic showing embryos ablated at non-migratory stages, and neural crest migration at stage 23 (migratory)[7]. **b**, Lateral views of control and posterior-ablated embryos hybridized with a probe against *snail2*. Scale bar, 250 µm. **c**, **d**, Targeted injections into the head mesoderm. **c**, Cartoon shows that injecting two dorso-vegetal blastomeres targets the head mesoderm. **d**, Confocal images of targeted injections showing nuclear RFP expression in the mesoderm. Scale bar, 100 µm. **e**–**g**, Compression controls. **e**, Images of embryos being compressed from non-migratory to pre-migratory stages with a 90-µm bead attached to an AFM cantilever (bead, red circumference). Neural plate border highlighted with red lines and neural crest position shown in light blue; stages as labelled. **f**, Magnitude maps from a particle image velocity (PIV) analysis indicates the *x*–*y* extent of the deformation induced by AFM indentation. **g**, Maximum projection of a cross-section showing the *z*-deformation generated by the AFM compression. Epidermis (ep) and mesoderm (ms) are being deformed, mean maximum indentation depth (δ) is provided (nuclei are visible in the bead region owing to a maximum projection effect). **h**, **i**, Result of extrinsic stress experiments. **h**, Dorsal view (anterior to top) of embryos hybridized with a probe against *snail2*; black arrows point to migrating neural crest. **i**, Percentage of embryos displaying neural crest migration. Histograms show mean, error bars represent s.e.m.; $n = 13$ embryos; two-tailed *t*-test, $^{**}P = 0.001$, CI = 95%. All experiments repeated three times. Scale bars: 200 µm (**e**, **f**), 100 µm (**g**), 200 µm (**h**).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



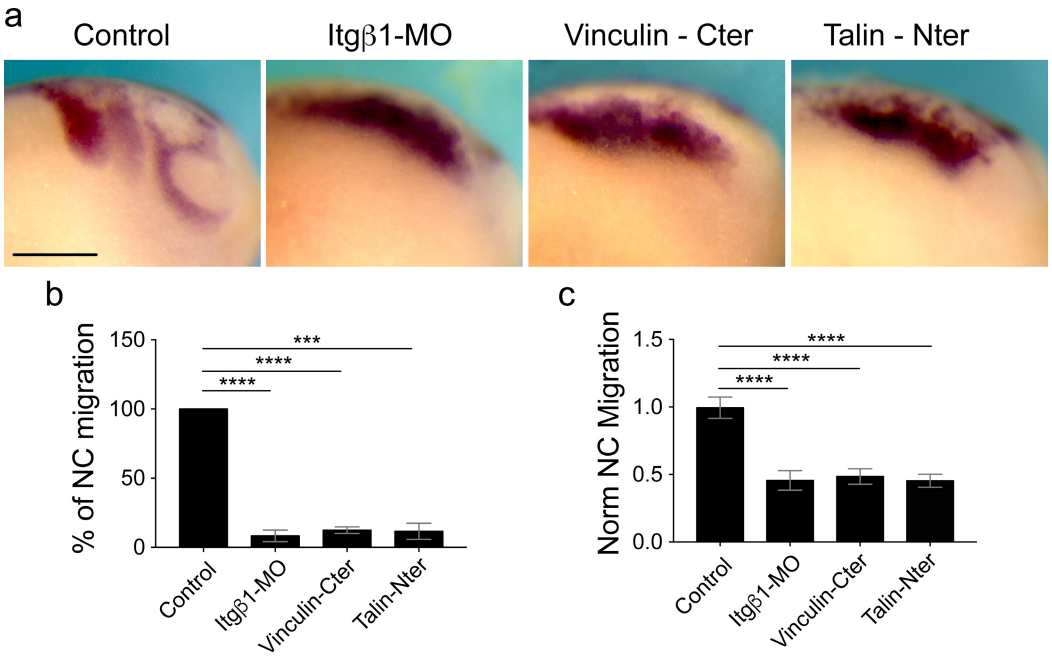
Extended Data Figure 5 | Mesodermal stiffening promotes neural crest CCM in a non-autonomous manner *in vivo*. **a**–**d**, Non-autonomous effect of mesodermal myosin manipulations on neural crest migration *in vivo*. **a**, Wild-type labelled pre-migratory neural crest (cyan) grafted into pre-migratory wild-type, *myl9*-MO or CA-MYPT-injected hosts (mesoderm in red). **b**, Representative embryos showing normal neural crest migration in wild-type hosts and inhibited neural crest migration in *myl9*-MO or CA-MYPT hosts. Scale bar, 200 μm. **c**, **d**, Normalized neural crest migration (**c**, $n = 22$ embryos) and percentage of embryos displaying neural crest migration (**d**, $n = 38$ embryos) (**c**, **d**; one-way ANOVA, $P < 0.0001$; two-tailed *t*-test, $^{***}P < 0.0006$, $^{****}P < 0.0001$, CI = 95%). Histograms in **c**, **d** and **g** show mean, error bars represent s.d. (**c**) or s.e.m. (**d**, **g**). **e**–**g**, Effect of premature mesodermal stiffening on neural crest migration *in vivo*. **e**, Wild-type labelled pre-migratory neural crest grafted into non-migratory wild-type or CA-MLC hosts. **f**, Representative embryos showing premature neural crest migration in CA-MLC hosts, whereas no migration is observed in control embryos at this stage. Scale bar, 100 μm. **g**, Percentage of embryos displaying neural crest migration. Two-tailed *t*-test, $^{****}P < 0.0001$, CI = 95%, $n = 20$ embryos. **h**–**k**, Non-autonomous effect of DshDEP^{+} on neural crest migration. **h**, Wild-type labelled pre-migratory neural crest (cyan) was grafted into pre-migratory wild-type or DshDEP^{+} hosts. **i**, Representative examples showing neural crest migration in wild-type hosts and inhibited neural crest migration in DshDEP^{+}-injected hosts (mesoderm in red). Scale bar, 150 μm. Normalized neural crest migration (**j**) and percentage of embryos displaying neural crest migration (**k**) (**j**, **k**; two-tailed *t*-test, $^{***}P < 0.0002$, $^{****}P < 0.0001$, CI = 95%, $n = 14$ embryos). Histograms in **j** and **k** show mean, error bars represent s.d. (**j**) and s.e.m. (**k**). **b**, **f** and **i** are representative examples of three independent experiments.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



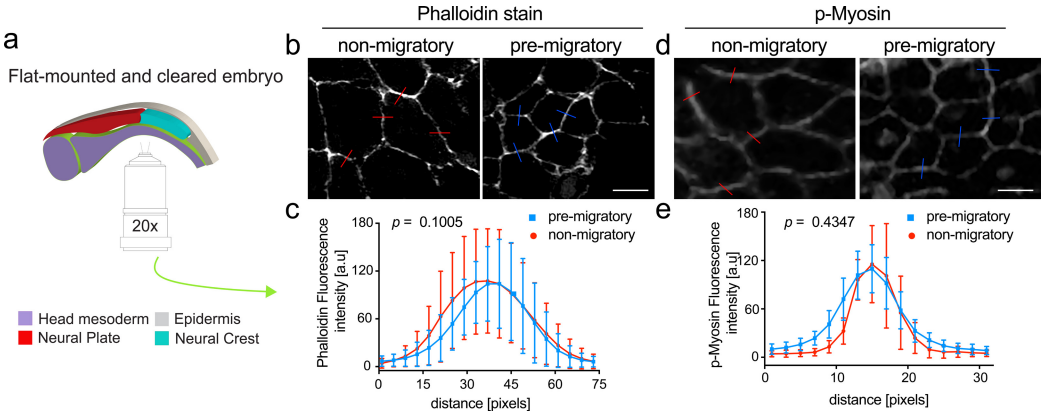
Extended Data Figure 6 | Head mesoderm manipulations did not affect fibronectin deposition and organization under the neural crest. As it has been previously shown that inhibition of myosin and PCP can affect fibronectin deposition and organization in the dorsal mesoderm[27,28,42]; we analysed the effect of these treatments on the lateral mesoderm underlying the premigratory neural crest. Our results show no effect of inhibition of myosin or PCP on fibronectin deposition or organization in the lateral mesoderm, indicating that the effect of our treatments on neural crest migration are not due to an indirect effect on fibronectin, but rather a consequence of affecting mesoderm stiffness. **a**–**c**, Transverse sections showing fibronectin expression with different treatments used to manipulate the apparent elasticity of head mesoderm. **a**, Schematic showing the area presented in **b** and **c**. **b**, **c**, Immunostaining against fibronectin (green) and DAPI stain (blue) showing the fibronectin signal underneath the neural crest, treatments as labelled in each panel. Scale bar, 30 μm. **d**–**g**, Effect of mesoderm manipulation on fibronectin assembly. **d**, Dorsal view of a flat-mounted embryo. White square indicates the area shown in **e** and **f**. **e**, Immunostaining against fibronectin (green) in non-migratory embryos (stage 13), showing fibronectin assembly underneath the neural crest, treatments as indicated. **f**, Immunostaining against fibronectin (green) in pre-migratory embryos (stage 20) showing fibronectin assembly underneath the neural crest, treatments as indicated. Control to show that inhibition of PCP signalling (DshDEP^{+}) affects fibronectin assembly in dorsal mesoderm[42] (**g**) but not lateral mesoderm (**f**). Scale bars, 30 μm. **b**, **c** and **e**–**g** are representative examples of three independent experiments.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



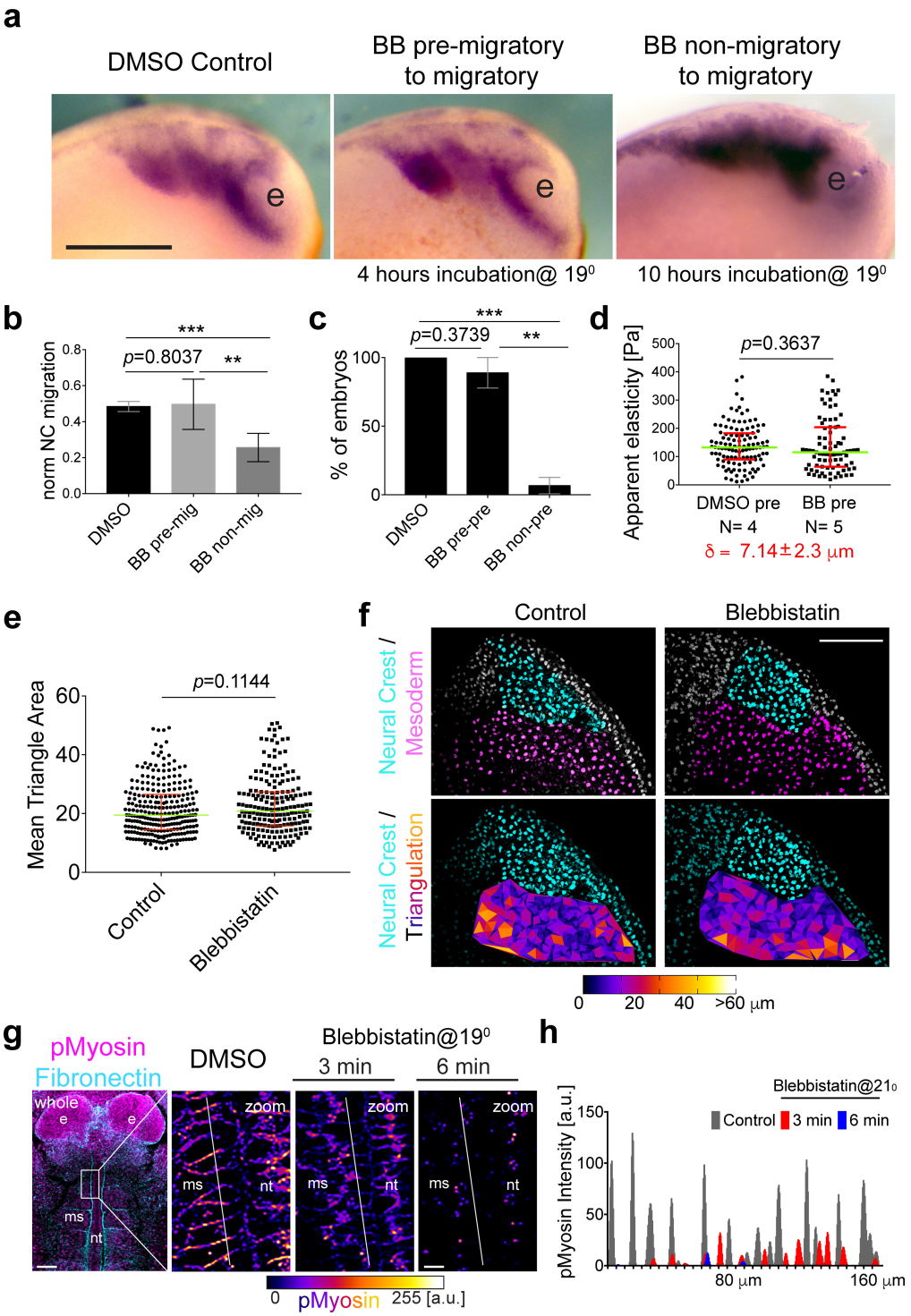
Extended Data Figure 7 | Neural crest senses mesoderm stiffening by using an integrin–vinculin–talin-mediated mechanoresponsive mechanism. a–c, Analysis of neural crest migration after integrin β1, vinculin or talin knockdown. **a**, Lateral views of embryos hybridized with a probe against *snail2*. Control embryos migrate normally from dorsal to ventral. However, neural crest migration is drastically inhibited in embryos treated with itgβ1-MO or the dominant-negative mutants vinculin-Cter and talin-Nter. Embryos in **a** are representative examples of three independent experiments. **b**, **c**, Quantification of neural crest migration showing percentage of embryos (**b**) and normalized neural crest migrated distances (**c**). Histograms in **b** and **c** show mean, error bars represent s.e.m. (**b**) and s.d. (**c**); one-way ANOVA, $P < 0.0001$; two-tailed *t*-test, $^{***}P < 0.0002$ $^{****}P < 0.0001$, CI = 95%; $n = 39$ embryos. Scale bar, 200 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



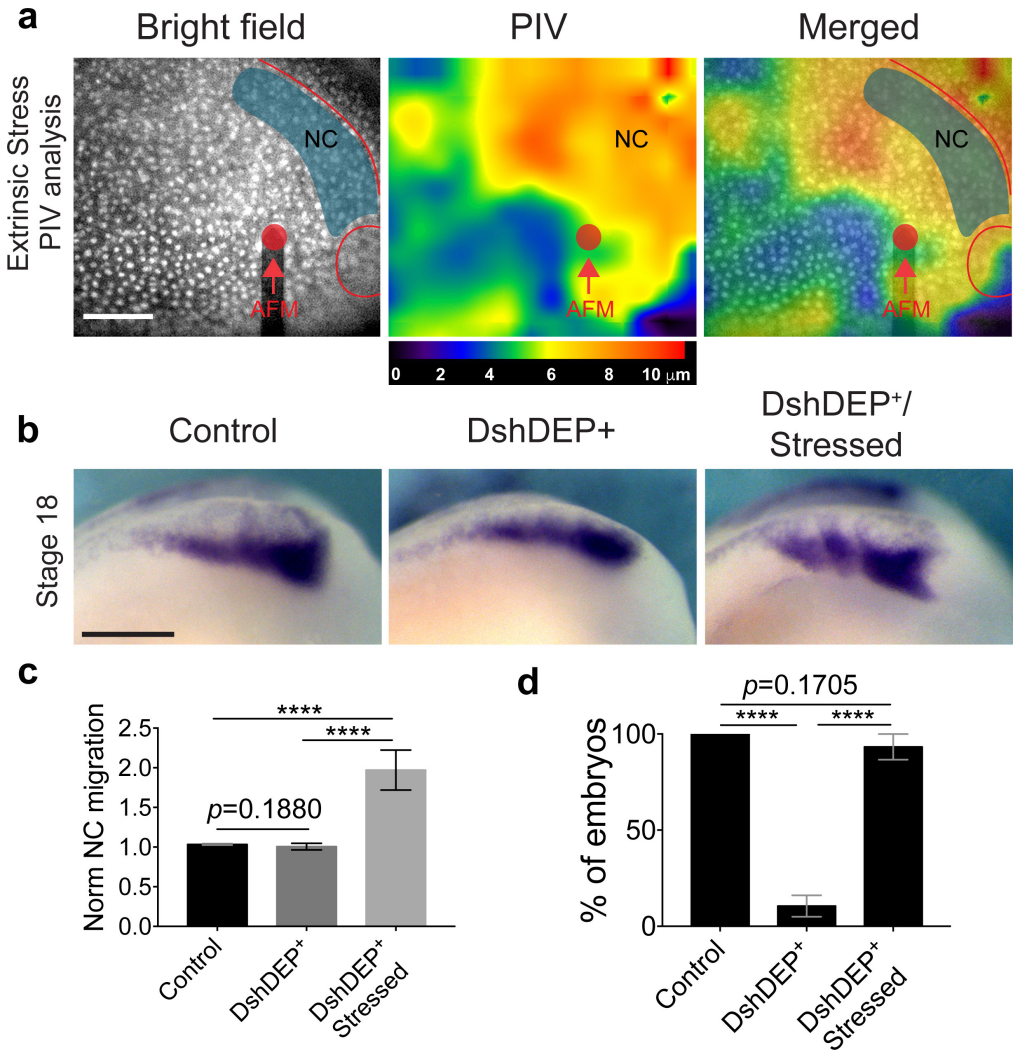
Extended Data Figure 8 | Actomyosin expression in the head mesoderm does not change from non- to pre-migratory stages. a–e, Analysis of actin and phospho-myosin expression under the neural crest. As actomyosin contractility has been proposed to contribute to paraxial mesoderm stiffness in *Xenopus*[13], we assessed actomyosin activity in the head mesoderm, the substrate of the neural crest. Note that the expression levels of either actin or phospho-myosin remained unchanged at pre-migratory stages compared to non-migratory stages. **a**, Schematic showing the imaging method used to access the head mesoderm under the neural crest. **b**, Phalloidin stain was used to label actin on embryos at non- and pre-migratory stages. **d**, Immunostaining against phospho-myosin on embryos at non- and pre-migratory stages. **c**, **e**, Quantification of fluorescence intensity across the membrane in **b** and **d**, respectively, represented by a continuous line. Points show mean and error bars represent s.d.; *n*, number of measurements per data (time) point; $n = 16$ (**c**), $n = 15$ (**e**); from three independent experiments; non-significant differences were observed, *P* value is indicated in each chart (two-tailed *t*-test, CI = 95%). Scale bars, 20 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 9 | Myosin contractility is not required to maintain mesodermal stiffness. a–f, Blebbistatin incubation experiments. To functionally test the effect of actomyosin contractility on mesoderm stiffness, we applied blebbistatin to pre-migratory embryos. Note that blebbistatin treatment at this stage did not affect mesoderm stiffness, cell density or neural crest migration, in spite of a marked reduction in myosin phosphorylation. **a**, Lateral views of embryos hybridized with a probe against *snail2*, showing no differences in neural crest migration after 4 h of incubation with DMSO or blebbistatin (BB). Embryos were incubated as indicated. Scale bar, 150 μm. Normalized neural crest migration (**b**) and percentage of embryos (**c**). In **b** and **c**, histograms show mean and error bars represent s.d. (**b**) and s.e.m. (**c**); one-way ANOVA, $P < 0.0001$; two-tailed *t*-test $^{**}P < 0.008$, $^{***}P < 0.002$, CI = 95%, $n = 19$ embryos. **d**, iAFM measurement direct on mesoderm. Spread of data plotted for each condition; green lines show median, red whiskers represent the interquartile range (two-tailed Mann–Whitney *U*-test, *P* value as indicated, CI = 95%; *n* = number of AFM indentations, $n = 116$ (DMSO), $n = 204$ (BB_pre). **e**, **f**, As blebbistatin treatment could potentially affect cell–cell adhesion, thus promoting cell dispersion, we measured its effect on cell dispersion by calculating the distance between neighbouring mesoderm cells. **e**, Chart comparing the mean area of the triangle formed by each mesoderm cell with its neighbours, calculated by Delaunay triangulation (Mann–Whitney *U*-test, *P* value is shown). *n* = number of embryos, δ = mean maximum indentation depth. **f**, Images showing representative neural crest and mesoderm cells (treated as indicated) and triangles from which the areas in **e** were calculated. Scale bar, 200 μm. **g**, Immunostaining for phospho-myosin; dorsal view of a flat-mounted embryo and colour-coded intensity of zoom images at the indicated time points are shown (scale in arbitrary units, a.u.). Scale bars, 100 μm (whole), 15 μm (zoom). **h**, Fluorescence intensity in a.u. across the white line in the zoom panels in **g**. All experiments repeated at least three times. Images in **a**, **f** and **g** are representative of three independent experiments.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | Extrinsic compression modifies the environment of PCP-depleted embryos and triggers premature neural crest CCM *in vivo*. **a**, Magnitude maps from a particle image velocity (PIV) analysis showing the *x*–*y* extent of the stress field generated by the extrinsic compression experiment. Neural plate border is indicated by red lines, neural crest is shown in light blue. Scale bar, 200 μm. **b**, **c**, Extrinsic stress-induced premature neural crest migration. **b**, Lateral views of embryos hybridized with a probe against *snail2*. Note that premature neural crest migration is observed in the DshDEP^{+}/stressed panel. Normalized neural crest migration (**c**) and percentage of embryos displaying each phenotype (**d**). Histograms show mean, error bars show s.d.(**c**) and s.e.m. (**d**); one-way ANOVA, $P < 0.0001$; two-tailed *t*-test, $^{****}P < 0.0001$, CI = 95%, $n = 18$ embryos. **a** and **b** show representative images of three independent experiments. Scale bar, 150 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature25506

Peptidoglycan synthesis drives an FtsZ–treadmilling–independent step of cytokinesis

João M. Monteiro[1]*, Ana R. Pereira[1]*, Nathalie T. Reichmann[1], Bruno M. Saraiva[1], Pedro B. Fernandes[1], Helena Veiga[1], Andreia C. Tavares[1], Margarida Santos[1], Maria T. Ferreira[1], Vânia Macário[1], Michael S. VanNieuwenhze[2], Sérgio R. Filipe[3] & Mariana G. Pinho[1]

Peptidoglycan is the main component of the bacterial wall and protects cells from the mechanical stress that results from high intracellular turgor. Peptidoglycan biosynthesis is very similar in all bacteria; bacterial shapes are therefore mainly determined by the spatial and temporal regulation of peptidoglycan synthesis rather than by the chemical composition of peptidoglycan. The form of rod-shaped bacteria, such as *Bacillus subtilis* or *Escherichia coli*, is generated by the action of two peptidoglycan synthesis machineries that act at the septum and at the lateral wall in processes coordinated by the cytoskeletal proteins FtsZ and MreB, respectively[1,2]. The tubulin homologue FtsZ is the first protein recruited to the division site, where it assembles in filaments—forming the Z ring—that undergo treadmilling and recruit later divisome proteins[3,4]. The rate of treadmilling in *B. subtilis* controls the rates of both peptidoglycan synthesis and cell division[3]. The actin homologue MreB forms discrete patches that move circumferentially around the cell in tracks perpendicular to the long axis of the cell, and organize the insertion of new cell wall during elongation[5,6]. Cocci such as *Staphylococcus aureus* possess only one type of peptidoglycan synthesis machinery[7,8], which is diverted from the cell periphery to the septum in preparation for division[9]. The molecular cue that coordinates this transition has remained elusive. Here we investigate the localization of *S. aureus* peptidoglycan biosynthesis proteins and show that the recruitment of the putative lipid II flippase MurJ to the septum, by the DivIB–DivIC–FtsL complex, drives peptidoglycan incorporation to the midcell. MurJ recruitment corresponds to a turning point in cytokinesis, which is slow and dependent on FtsZ treadmilling before MurJ arrival but becomes faster and independent of FtsZ treadmilling after peptidoglycan synthesis activity is directed to the septum, where it provides additional force for cell envelope constriction.

In cocci, one possible molecular cue determining the shift of peptidoglycan (PG) synthesis from the cell periphery to the septum could be the recruitment to the midcell of a key PG biosynthesis protein, concomitant with divisome assembly. We therefore investigated the localization of most of the *S. aureus* PG synthesis proteins in the background of the methicillin-resistant *S. aureus* (MRSA) strain COL (Fig. 1 and Supplementary Table 1). All fluorescent fusions were functional (Supplementary Table 1) and expressed from their native locus under the control of their native promoter as the sole copy of the gene in the cell, with the exception of MraY–sGFP. As expected MurB, MurD and MurF fusions, which act on cytoplasmic PG precursors, showed cytoplasmic localization (Fig. 1b). Consistent with their substrate localization, the remaining fusions localized to the membrane; these fusions included the FemXAB proteins, which do not have canonical membrane-targeting domains[10]. Because most PG synthesis activity occurs at the septum[9], we were expecting membrane-associated PG synthesis enzymes to be highly enriched in the septal region of dividing cells. However, MraY, MurG and the FemXAB proteins were evenly distributed throughout the membrane (including the septum), which suggests that the key step for spatial regulation of PG synthesis was not the synthesis of lipids I or II (Fig. 1b, c). MurJ, FtsW and PBP1 were the only septal proteins for which virtually no signal could be observed in the peripheral membrane during septum synthesis (Fig. 1b, c), and therefore represented good potential candidates as catalysts for the first step of PG synthesis that is diverted to the septal region. MurJ is a member of the multidrug–oligosaccharidyl-lipid–polysaccharide (MOP) exporter superfamily and has previously been suggested to be the lipid II flippase in *E. coli*[11]. In *S. aureus*, the essential gene *SAV1754* (also known as *SACOL1804*) has previously been reported to be a functional MurJ orthologue[12]. FtsW is a member of the sporulation, elongation, division and synthesis (SEDS) protein family, also suggested to be a lipid II flippase[13]. However, SEDS proteins have more recently been shown to be PG transglycosylases that probably function together with a cognate penicillin-binding protein (PBP) during PG polymerization[14,15]. *S. aureus* encodes two SEDS proteins, SACOL1122 and SACOL2075, which are similar to the *B. subtilis* proteins FtsW and RodA, respectively. PBP1 is a transpeptidase that crosslinks PG glycan strands[16].

To clarify which protein or set of proteins was responsible for directing PG synthesis to the septum, newly synthesized PG was labelled with the fluorescent D-amino acid HCC-amino-D-alanine (HADA) which is specifically incorporated into PG[17], in an *S. aureus* strain expressing both FtsW–mCherry (which co-localizes with PBP1, see later) and MurJ–sGFP. HADA incorporation appeared to co-localize with both proteins in cells in phase 1 of the cell cycle and in most phase 2 cells (Fig. 2a; see Fig. 1d for cell-cycle phases). However, MurJ and HADA septal co-localization was more frequent than FtsW and HADA septal co-localization (88% versus 70% of cells, $n = 200$), as cells with septal FtsW but peripheral MurJ had peripheral HADA incorporation (see asterisks in Fig. 2a). This suggested that septal PG synthesis was dependent on the presence of MurJ. If this was the case, preventing MurJ recruitment to the midcell should abolish septal PG synthesis. We therefore investigated the mechanism of MurJ localization so that we could selectively prevent its septal recruitment while maintaining correct FtsW and PBP1 septal localization. For this purpose, we determined the timing of MurJ arrival to the septum, because the localization of divisome proteins is often dependent on the presence of earlier localizing proteins[18,19].

In *B. subtilis*, divisome assembly is a two-step process: proteins such as FtsA, ZapA and EzrA arrive very early and concomitantly with FtsZ, and are followed after a time delay by a second group of proteins including the DivIB–DivIC–FtsL sub-complex[20]. We therefore compared *S. aureus* PBP1, FtsW and MurJ localization with that of

[1]Bacterial Cell Biology, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal. [2]Department of Chemistry, Indiana University, Bloomington, Indiana, USA. [3]Instituto de Tecnologia Química e Biológica António Xavier and Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Oeiras, Portugal.
*These authors contributed equally to this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

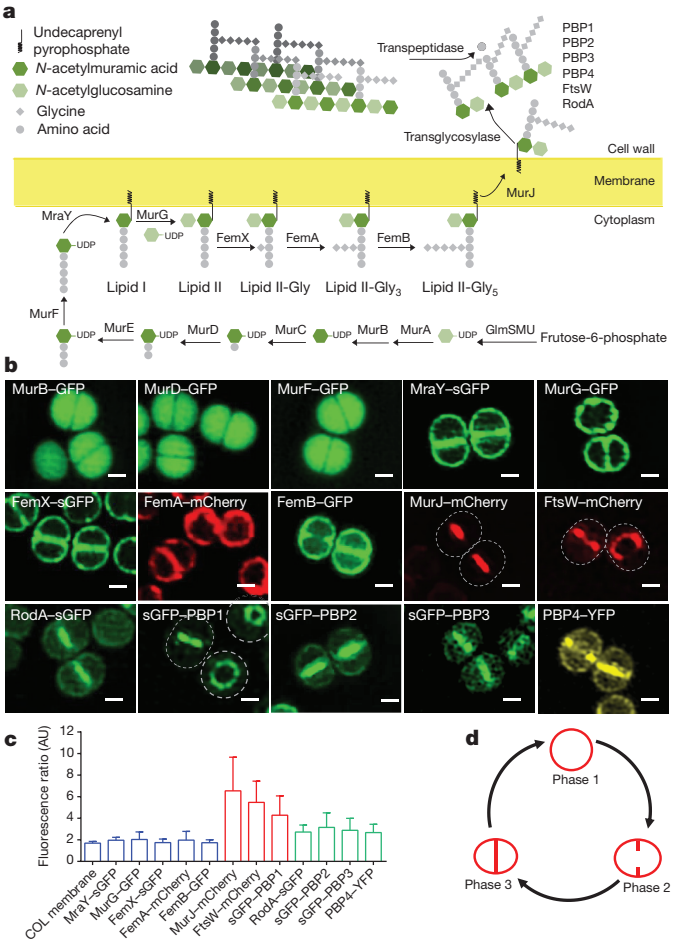


Figure 1 | Localization of PG synthesis proteins during division of *S. aureus*. **a**, Schematic of the peptidoglycan synthesis pathway in *S. aureus*. UDP, uridine diphosphate. **b**, Structured illumination microscopy (SIM) images of *S. aureus* cells expressing fluorescent derivatives of PG synthesis proteins. Scale bars, 0.5 μm. Images are representative of three biological replicates. **c**, Ratios between fluorescence signal at the septum and the peripheral membrane, measured in cells with a complete septum (phase 3). Blue, membrane proteins with fluorescence ratio of approximately 2, similar to Nile Red staining of COL membrane (expected as the septum contains two membranes versus one in the cell periphery). Green, septal enriched proteins with $2.5 <$ fluorescence ratio < 3.5. Red, septal proteins with fluorescence ratio > 4. Data are represented as column graphs in which the height of the column is the mean and whiskers are s.d. From left to right, $n = 439$, 533, 516, 513, 512, 622, 503, 517, 503, 689, 1,321, 488 and 516 cells. AU, arbitrary unit. **d**, Schematic of the *S. aureus* cell cycle. Phase 1 cells have not initiated septum synthesis, phase 2 cells are undergoing septum synthesis and phase 3 cells have a complete septum undergoing maturation in preparation for splitting.

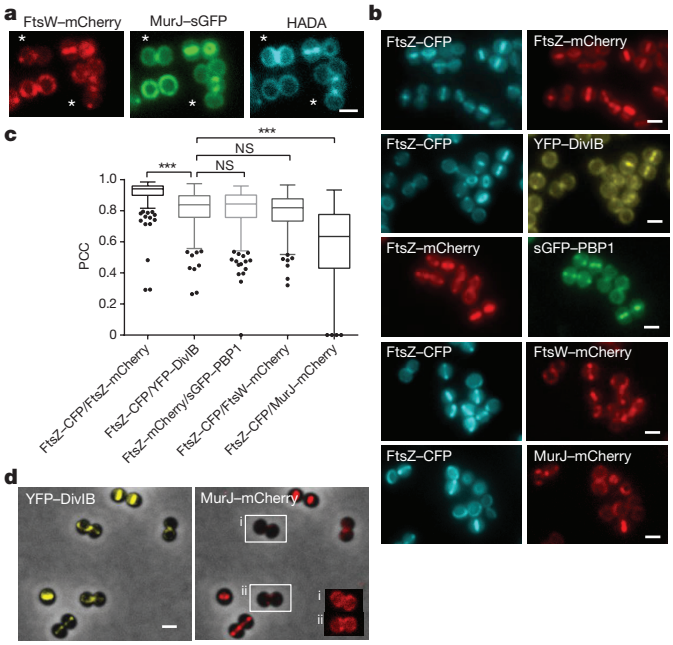


Figure 2 | Hierarchy of the arrival of key PG synthesis proteins to the divisome. **a**, PG synthesis in strain ColWJ was labelled with HADA. Asterisks show cells with septal FtsW, peripheral MurJ and HADA labelling. **b**, Fluorescence microscopy images of strains ColZZ (positive control), ColZIB, ColP1Z, ColWZ and ColJZ (strains from top to bottom). **c**, PCC for pairs of fluorescent proteins expressed in the same strain. A PCC of 1 would indicate perfect co-localization; negative PCC values are represented as 0. From left to right, $n = 312$, 391, 337, 330 and 334 cells. Data are represented as box-and-whisker plots in which boxes correspond to the first-to-third quartiles, lines inside the boxes indicate the median, and the ends of whiskers and outliers follow a Tukey representation. Statistical analysis was performed using a two-sided Mann–Whitney *U* test. $***P < 0.001$; first versus second sample, $P = 3.65 \times 10^{-45}$; second versus fifth sample, $P = 1.13 \times 10^{-45}$; NS, not significant. **d**, Fluorescence microscopy images of strain ColJIB overlayed with phase contrast images. Cells with septal YFP–DivIB localization, but in which MurJ–mCherry is still not at the septum (peripheral), are highlighted in insets. In these insets, the brightness of the fluorescence signal was increased to enable visualization of MurJ localization. Scale bars, 1 μm. Images in **a**, **b** and **d** are representative of three biological replicates.

the early divisome protein FtsZ and the later divisome protein DivIB (Fig. 2b). Co-localization between each protein and FtsZ was determined by measuring Pearson's correlation coefficient (PCC) for the fluorescence signals in the two channels in cells that showed FtsZ midcell localization. We reasoned that proteins arriving to the septum simultaneously with FtsZ should have a PCC close to 1 and that this value should decrease for later divisome proteins. We constructed an *S. aureus* strain that co-expressed FtsZ–CFP and FtsZ–mCherry as a positive control. Co-localization results indicated that DivIB arrives at the divisome later than FtsZ, as expected, and at approximately the same time as PBP1 and FtsW, whereas MurJ arrives later than DivIB, PBP1 and FtsW (Fig. 2c); this result is unlikely to be affected by the nature of the fluorescent tags (Extended Data Fig. 1). Consistent with these timings, in 20% of the cells ($n = 200$) of an *S. aureus* strain that expressed both YFP–DivIB and MurJ–mCherry, YFP–DivIB was already localized at the septum whereas MurJ–mCherry had not yet arrived (insets in Fig. 2d). This raised the possibility that MurJ septal recruitment is dependent on the presence of the DivIB–DivIC–FtsL sub-complex. We therefore depleted expression of each of these three proteins using antisense RNA fragments[21] (Supplementary Table 2) and assessed the efficiency of protein depletion by western blot analysis and by the increase in cell volume that resulted from divisome inhibition (Extended Data Fig. 2). Depletion of DivIB, DivIC or FtsL reduced septal co-localization of FtsZ–CFP and MurJ–mCherry (Fig. 3a, b), but not of FtsZ and either FtsW or PBP1 (Extended Data Fig. 3), consistent with their earlier recruitment to the divisome.

Having developed a tool to specifically delocalize MurJ by depleting FtsL while at the same time maintaining correct PBP1 and FtsW localization, we were able to determine that new PG, labelled by a short pulse of HADA, was only incorporated at the midcell if MurJ–sGFP was also present at the septum (Fig. 3c asterisks, 3d). Finally, we showed that inhibiting MurJ activity using DMPI, a MurJ inhibitor[12] that does not prevent the recruitment of MurJ to the divisome (Extended Data Fig. 4a), markedly reduced HADA incorporation (that is, PG synthesis) at the septum (Extended Data Fig. 4b–e).

Taken together, these data indicate that recruitment of MurJ to the divisome by the DivIB–DivIC–FtsL complex is likely to be the molecular cue that directs PG synthesis specifically to the septum

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

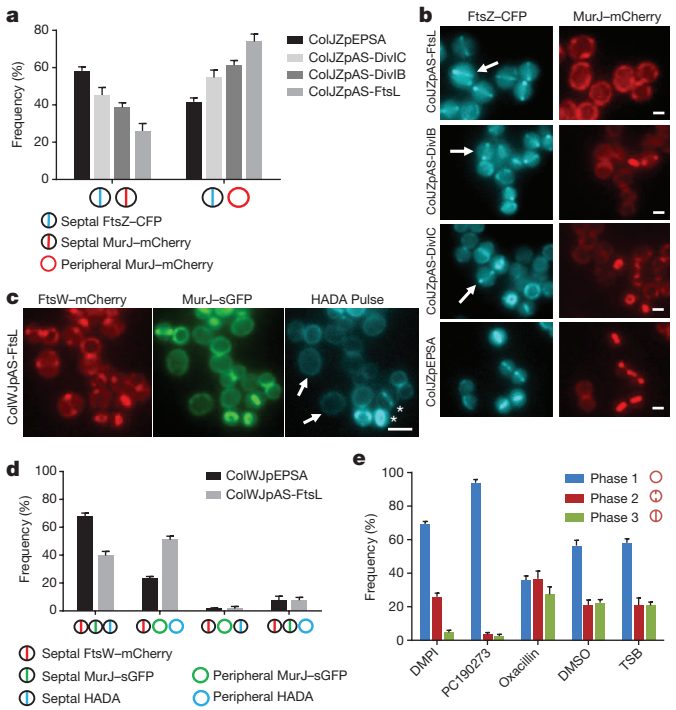


Figure 3 | Recruitment of MurJ to the divisome is required for PG incorporation at the septum. a, **b**, FtsZ–CFP and MurJ−mCherry co-localization (**a**) and microscopy images (**b**) of cells expressing antisense RNA against *ftsL*, *divIB* or *divIC*, or carrying vector pEPSA ($n = 100$ cells for each strain, in each of three biological replicates). Arrows show large cells with septal FtsZ–CFP, but delocalized MurJ–mCherry. Scale bars, 1 μm. **c**, In FtsL-depleted cells, septal incorporation of HADA was observed only in cells with septal MurJ–sGFP (asterisks). Septal FtsW–mCherry was not sufficient to ensure septal incorporation of HADA (arrows). Scale bar, 2 μm. **d**, Co-localization between FtsW–mCherry, MurJ–sGFP and HADA labelling in FtsL-depleted (ColWJpAS-FtsL) or control (ColWJpEPSA) cells with septal FtsW–mCherry ($n = 100$ cells for each strain, in each of three biological repeats). **e**, COL cells treated with DMPI, PC190723, oxacillin, dimethyl sulfoxide (DMSO) or tryptic soy broth (TSB) (mock-treated controls) for the duration of a cell cycle (see Extended Data Fig. 5) were classified into each cell-cycle phase ($n = 349$–870 cells for each condition, in each of the three biological repeats). Data in **a**, **d** and **e** are column graphs in which the height of the column represents the mean, and the whiskers are the s.d. Images in **b** and **c** are representative of three biological replicates.

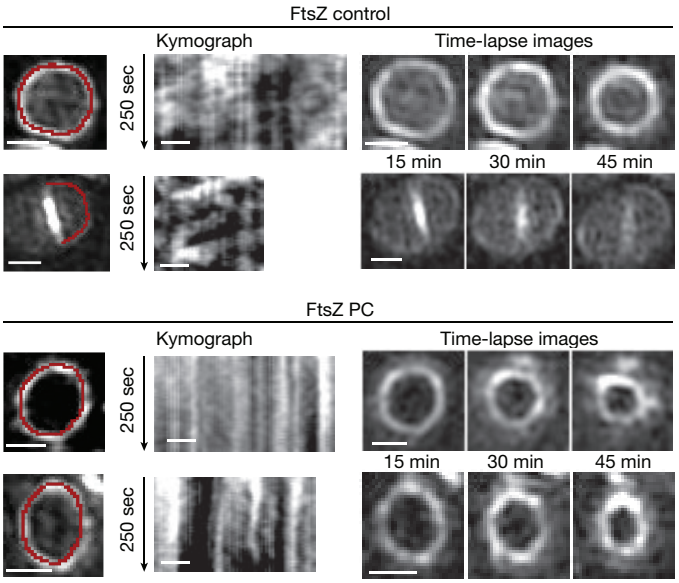


Figure 4 | PC190723 impairs FtsZ$^{55–56}$sGFP treadmilling in *S. aureus*. ColFtsZ$^{55–56}$sGFP cells were imaged by SIM every 5 s for 250 s, followed by three images taken 15 min apart, in the absence (FtsZ control) or presence (FtsZ PC) of PC190723. Kymographs were obtained by extracting fluorescence intensity values along the red line indicated in cells in the left panels, with each horizontal line of the kymograph corresponding to one time point. In the absence of PC190723 (upper panels) FtsZ movement is observed, producing diagonal lines in the kymograph. The addition of PC190723 (lower panels) abolished FtsZ movement, producing vertical lines in the kymograph that correspond to maintenance of signal at fixed locations, but did not impair Z-ring constriction (time-lapse images on the right). Data are representative of three biological replicates. Scale bars, 0.5 μm.

during division. Therefore, we would expect MurJ to be essential for the transition from phase 1 to phase 2 during the cell cycle (that is, for initiation of septum synthesis). We treated COL wild-type cells with DMPI for one cell cycle and characterized the distribution of cells in the three cell-cycle phases. For comparison we tested other inhibitors, namely PC190723[22], which targets FtsZ, and oxacillin, which inhibits PG transpeptidation catalysed by PBPs (Fig. 3e and Extended Data Fig. 5). Consistent with previous data[9], in the absence of inhibitors approximately half of the cells were in phase 1, with the other half split evenly between phase 2 and phase 3 (Fig. 3e). Inhibition of MurJ by DMPI resulted in the accumulation of phase 1 cells (70%, $n > 300$), which shows that MurJ is indeed crucial for entering phase 2. Phase 2 cells, which are synthesizing the septum, were also halted, as flipping the PG precursor is essential for PG synthesis to occur.

The most notable result was the almost complete absence of phase 2 cells in the presence of PC190723. This compound inhibits FtsZ treadmilling[3], which has been shown to control the rate of PG synthesis during septum formation in *B. subtilis* and therefore to control the rate of cell division[3]. Consequently, one would expect that the addition of PC190723 would prevent phase 2 cells that were halfway through the process of septum synthesis from completing this process. As this was not the case we considered whether the redirection of PG synthesis to the septum—which is driven by septal recruitment of MurJ—could provide the constrictive driving force for cytokinesis, as has previously been suggested for *E. coli*[23,24], and result in the closing of the septum independent of FtsZ treadmilling.

To visualize the closing of the septum and determine whether FtsZ treadmilling occurs in *S. aureus*, we introduced an FtsZ sandwich fusion to superfast GFP (sGFP), in which *sgfp* was inserted in the *ftsZ* coding sequence between codons 55 and 56 (FtsZ$^{55–56}$sGFP; see Methods) in the background of a strain that expressed native FtsZ, and then followed the dynamics of Z-ring formation and constriction. We were able to observe the movement of FtsZ$^{55–56}$sGFP, which as expected was inhibited by PC190723 (Fig. 4 and Supplementary Video 1). However, Z-ring constriction continued in many cells treated with PC190723, consistent with the fact that PC190723-treated cells could complete phase 2 of the cell cycle (Fig. 3e). When we followed Z-ring constriction in untreated cells we observed biphasic cytokinesis, with a first step immediately after Z-ring assembly during which the divisome barely constricts, followed by a second step with a higher rate of Z-ring constriction (Fig. 5a, Extended Data Fig. 6a and Supplementary Video 2). The addition of PC190723 blocked the constriction of large Z rings, presumably in the first step of cytokinesis, but not of smaller Z rings (Fig. 5a, b, Extended Data Fig. 6a and Supplementary Video 3). This shows that only Z rings in the first step of cytokinesis required treadmilling activity for constriction. To confirm these results, we used a functional, fluorescent derivative of the divisome protein EzrA as a proxy for FtsZ localization, because EzrA interacts directly with FtsZ[25,26]. Similar to our observations for FtsZ, EzrA treadmilling was inhibited by PC190723, EzrA rings underwent biphasic constriction and the second and faster step of cytokinesis was insensitive to PC190723 (Extended Data Fig. 7).

It is possible that the transition between the first and the second step of Z-ring constriction corresponds to the start of substantial PG

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

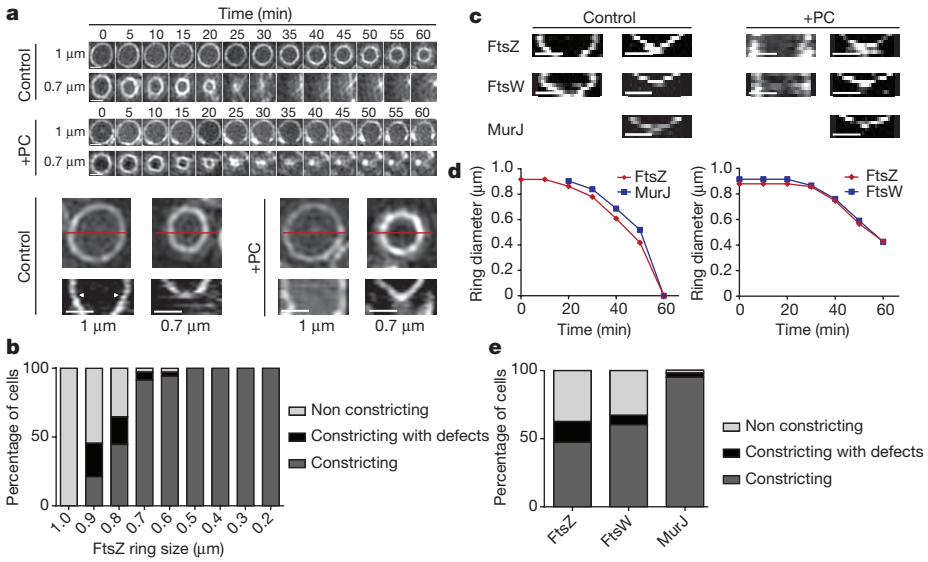


Figure 5 | PG synthesis provides force for septum constriction. a, Time-lapse images and corresponding kymographs of FtsZ$^{55–56}$sGFP rings in the absence (control) or presence (+PC) of the FtsZ inhibitor PC190723. Kymographs were obtained by drawing a line (example shown in red) across the cell for each time point. Arrowheads point to the transition from the first (no or slow constriction) to the second (fast constriction) step of cytokinesis. **b**, Percentage of FtsZ$^{55–56}$sGFP rings of different diameters that constricted in the presence of PC190723 ($n = 248$). **c**, Kymographs of FtsZ$^{55–56}$sGFP, FtsW–sGFP and MurJ–sGFP larger (left) and smaller (right) rings. PC190723 abolished only the constriction of larger rings. **d**, Graphs showing diameter of FtsZ–mCherry and MurJ–sGFP rings (left) or FtsZ–mCherry and FtsW–sGFP rings (right) in single cells during cytokinesis. **e**, Percentage of FtsZ$^{55–56}$sGFP, FtsW–sGFP and MurJ–sGFP rings that constricted in the presence of PC190723 (FtsZ, $n = 254$; FtsW, $n = 176$; MurJ = 188). Data in **a** and **c** are representative of three biological replicates. Scale bars, 0.5 μm.

synthesis activity that results from MurJ recruitment. Consistent with this hypothesis, the divisome rings that contained MurJ did not display the first step of constriction and were insensitive to PC190723 (Fig. 5c, e and Extended Data Figs 6b, 8). Furthermore, the arrival of MurJ at the septum coincided with the initiation of fast constriction (Fig. 5d and Extended Data Fig. 9a). By contrast, rings that contained the earlier divisome protein FtsW paralleled the biphasic behaviour of the Z ring (Fig. 5c, d, Extended Data Figs 8, 9b) and were susceptible to inhibition by PC190723, presumably during the initial stages of cytokinesis (Fig. 5e). The FtsZ inhibitor PC190723 blocked the constriction of Z rings only at initial stages, and the addition of the MurJ inhibitor DMPI blocked ring constriction at all stages (Extended Data Fig. 6a).

We propose a model (Extended Data Fig. 10) in which the *S. aureus* PG synthesis machinery continuously incorporates PG at the periphery of the cell during initial stages of the cell cycle. In preparation for division, and following Z-ring assembly, the DivIB–DivIC–FtsL complex recruits MurJ to the divisome, which ensures that translocation of lipid II occurs exclusively at the midcell. Substrate affinity[27] then diverts the major PG synthase, PBP2, from the periphery to the midcell where together with other PG synthesis enzymes it incorporates lipid II into the growing PG network. This mechanism means that an additional dedicated multi-protein machinery is unnecessary, and represents a mode of controlling PG synthesis in two different locations in the absence of an MreB-like cytoskeleton.

After the initiation of large-scale PG synthesis activity at the leading edge of the constricting septum that follows MurJ recruitment, the FtsZ inhibitor PC190723, which inhibits FtsZ treadmilling, no longer prevents cytokinesis. Nevertheless, FtsZ treadmilling is likely to have a role in the organization of septum synthesis, because approximately 15% of the Z rings that were able to constrict in the presence of PC190723 did so defectively, in a similar fashion to *E. coli* FtsZ mutants that are impaired in GTPase activity[4].

Our data may reconcile two models of the origin of the force required for cytokinesis to occur. In one model, this force has been proposed to be derived from FtsZ, either from the chemical energy of GTP hydrolysis that could promote bending of the FtsZ polymers or from the propensity for FtsZ filaments to bundle, which could result in condensation of the Z ring[24,28]. Alternatively, PG synthesis has previously been suggested to provide the force for cytokinesis[23,24]. We propose that cytokinesis occurs in two steps: an initial slow step, which is dependent on FtsZ treadmilling, for which FtsZ may be the driving force and that may be responsible for the initial invagination of the cell membrane, followed by a second and faster step for which PG synthesis provides the driving force.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 July 2017; accepted 9 January 2018.
Published online 14 February 2018.

1. Chastanet, A. & Carballido-Lopez, R. The actin-like MreB proteins in *Bacillus subtilis*: a new turn. *Front. Biosci. (Schol. Ed.)* **4,** 1582–1606 (2012).
2. Young, K. D. Bacterial shape: two-dimensional questions and possibilities. *Annu. Rev. Microbiol.* **64,** 223–240 (2010).
3. Bisson-Filho, A. W. *et al.* Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. *Science* **355,** 739–743 (2017).
4. Yang, X. *et al.* GTPase activity-coupled treadmilling of the bacterial tubulin FtsZ organizes septal cell wall synthesis. *Science* **355,** 744–747 (2017).
5. Domínguez-Escobar, J. *et al.* Processive movement of MreB-associated cell wall biosynthetic complexes in bacteria. *Science* **333,** 225–228 (2011).
6. Garner, E. C. *et al.* Coupled, circumferential motions of the cell wall synthesis machinery and MreB filaments in *B. subtilis*. *Science* **333,** 222–225 (2011).
7. Pinho, M. G. & Errington, J. Dispersed mode of *Staphylococcus aureus* cell wall synthesis in the absence of the division machinery. *Mol. Microbiol.* **50,** 871–881 (2003).
8. Pinho, M. G., Kjos, M. & Veening, J.-W. How to get (a)round: mechanisms controlling growth and division of coccoid bacteria. *Nat. Rev. Microbiol.* **11,** 601–614 (2013).
9. Monteiro, J. M. *et al.* Cell shape dynamics during the staphylococcal cell cycle. *Nat. Commun.* **6,** 8055 (2015).
10. Benson, T. E. *et al.* X-ray crystal structure of *Staphylococcus aureus* FemA. *Structure* **10,** 1107–1115 (2002).
11. Sham, L. T. *et al.* Bacterial cell wall. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science* **345,** 220–222 (2014).
12. Huber, J. *et al.* Chemical genetic identification of peptidoglycan inhibitors potentiating carbapenem activity against methicillin-resistant *Staphylococcus aureus*. *Chem. Biol.* **16,** 837–848 (2009).
13. Mohammadi, T. *et al.* Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. *EMBO J.* **30,** 1425–1432 (2011).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

14. Meeske, A. J. *et al.* SEDS proteins are a widespread family of bacterial cell wall polymerases. *Nature* **537,** 634–638 (2016).
15. Emami, K. *et al.* RodA as the missing glycosyltransferase in *Bacillus subtilis* and antibiotic discovery for the peptidoglycan polymerase pathway. *Nat. Microbiol.* **2,** 16253 (2017).
16. Wada, A. & Watanabe, H. Penicillin-binding protein 1 of *Staphylococcus aureus* is essential for growth. *J. Bacteriol.* **180,** 2759–2765 (1998).
17. Kuru, E. *et al. In situ* probing of newly synthesized peptidoglycan in live bacteria with fluorescent D-amino acids. *Angew. Chem. Int. Edn Engl.* **51,** 12519–12523 (2012).
18. Goehring, N. W., Gonzalez, M. D. & Beckwith, J. Premature targeting of cell division proteins to midcell reveals hierarchies of protein interactions involved in divisome assembly. *Mol. Microbiol.* **61,** 33–45 (2006).
19. Errington, J., Daniel, R. A. & Scheffers, D. J. Cytokinesis in bacteria. *Microbiol. Mol. Biol. Rev.* **67,** 52–65 (2003).
20. Gamba, P., Veening, J. W., Saunders, N. J., Hamoen, L. W. & Daniel, R. A. Two-step assembly dynamics of the *Bacillus subtilis* divisome. *J. Bacteriol.* **191,** 4186–4194 (2009).
21. Donald, R. G. *et al.* A *Staphylococcus aureus* fitness test platform for mechanism-based profiling of antibacterial compounds. *Chem. Biol.* **16,** 826–836 (2009).
22. Haydon, D. J. *et al.* An inhibitor of FtsZ with potent and selective anti-staphylococcal activity. *Science* **321,** 1673–1675 (2008).
23. Coltharp, C., Buss, J., Plumer, T. M. & Xiao, J. Defining the rate-limiting processes of bacterial cytokinesis. *Proc. Natl Acad. Sci. USA* **113,** E1044–E1053 (2016).
24. Coltharp, C. & Xiao, J. Beyond force generation: why is a dynamic ring of FtsZ polymers essential for bacterial cytokinesis? *BioEssays* **39,** e201600179 (2017).
25. Son, S. H. & Lee, H. H. The N-terminal domain of EzrA binds to the C terminus of FtsZ to inhibit *Staphylococcus aureus* FtsZ polymerization. *Biochem. Biophys. Res. Commun.* **433,** 108–114 (2013).
26. Levin, P. A., Kurtser, I. G. & Grossman, A. D. Identification and characterization of a negative regulator of FtsZ ring formation in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **96,** 9642–9647 (1999).
27. Pinho, M. G. & Errington, J. Recruitment of penicillin-binding protein PBP2 to the division site of *Staphylococcus aureus* is dependent on its transpeptidation substrates. *Mol. Microbiol.* **55,** 799–807 (2005).
28. Li, Z., Trimble, M. J., Brun, Y. V. & Jensen, G. J. The structure of FtsZ filaments *in vivo* suggests a force-generating role in cell division. *EMBO J.* **26,** 4694–4708 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Sobral for the construction of pSG-murF plasmid; A. Jorge for the construction of the pSG-EzrA plasmid; T. Roemer for providing DMPI and strains AS-022 and AS-185; M. DeLisa for pTRC99a-P7; Richard Novick for pCN51; S. Foster for antibodies against DivIB and DivIC; A. Henriques for critical reading of the manuscript and L. Krippahl for help with image analysis tools. This study was funded by the European Research Council through grant ERC-2012-StG-310987 (to M.G.P.), by Project LISBOA-01-0145-FEDER-007660 Microbiologia Molecular, Estrutural e Celular (to ITQB-NOVA), by the National Institutes of Health through grant NIHGM113172 (to M.V.N.) and FCT fellowships SFRH/BD/71993/2010 (J.M.M.), SFRH/BD/86416/2012 (A.R.P.), SFRH/BPD/95031/2013 (N.T.R.), SFRH/BPD/87374/2012 (H.V.), SFRH/BD/52204/2013 (A.C.T.) and SFRH/BD/77849/2011 (M.T.F.).

Author Contributions J.M.M., A.R.P., N.T.R. and M.G.P. designed the research. J.M.M. and A.R.P. performed all experiments with the exception of the effect of antibiotics on the cell cycle which was performed by P.B.F. B.M.S. developed software for image analysis. J.M.M., A.R.P., N.T.R., P.B.F., H.V., A.C.T, M.S., M.T.F. and V.M. constructed strains. M.S.V. contributed new reagents (HADA). J.M.M., A.R.P., N.T.R., S.R.F. and M.G.P. analysed the overall data. B.M.S. analysed microscopy data quantified by eHooke software. P.B.F. analysed cell-cycle data. J.M.M., A.R.P, N.T.R. and M.G.P. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.G.P. (mgpinho@itqb.unl.pt).

Reviewer Information *Nature* thanks J. Xiao and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessments.

Bacterial growth conditions. Strains and plasmids used in this study are listed in Supplementary Table 3. *S. aureus* strains were grown in tryptic soy broth (TSB, Difco) at 200 r.p.m with aeration at 37 °C or on tryptic soy agar (TSA, Difco) at 30 or 37 °C. *E. coli* strains were grown in Luria–Bertani broth (Difco) with aeration, or in Luria–Bertani agar (Difco), at 30 or 37 °C. When necessary, the antibiotics ampicillin (100 μg ml^{-1}), erythromycin (10 μg ml^{-1}), kanamycin (50 μg ml^{-1}), neomycin (50 μg ml^{-1}) or chloramphenicol (30 μg ml^{-1}) were added to the media. 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal, Apollo Scientific) was used at 100 μg ml^{-1}. Unless stated otherwise, isopropyl β-D-1-thiogalactopyranoside (IPTG, Apollo Scientific) was used at 0.1 mM to induce expression of constructs under the control of the Pspac promoter. Cadmium chloride (Sigma-Aldrich) was used at 0.1 μM when required, to induce expression of constructs under the control of the Pcad promoter.

Construction of *S. aureus* fluorescent strains. Cloning of fluorescent fusions in *S. aureus* was done using the following general strategy: plasmids were propagated in *E.coli* strains DC10B or DH5α and purified from overnight cultures supplemented with the relevant antibiotics. Plasmids were then introduced into electrocompetent *S. aureus* RN4220 cells as previously described[29] and transduced to COL using phage 80α[30]. Constructs were confirmed by PCR and by sequencing of the amplified fragment.

The ColMurB-GFP, ColMurD-GFP, ColMurF-GFP, ColFemB-GFP and ColpSGEzrA-GFP strains were constructed using the pSG5082 vector[31]. In brief, DNA fragments with approximately 500 bp spanning the 3′ ends (minus the stop codons) of the *murB*, *murD*, *murF*, *femB* and *ezrA* genes from COL were amplified using the primer pairs murBg_P1–murBg_P2, murDg_P1–murDg_P2, murFg_P1–murFg_P2, femBg_P1–femBg_P2 and ezrAP8Kpn–ezrAP9Xho, respectively (Supplementary Table 4). Fragments were digested with KpnI and XhoI restriction enzymes and cloned into pSG5082 upstream of and in frame with *gfpmutP2*, producing the plasmids pSG-murB, pSG-murD, pSG-murF, pSG-femB and pSG-ezrA. These plasmids were then electroporated to RN4220, where they integrated in the genome by a single homologous recombination event and subsequently transduced to COL. Resulting strains contained the corresponding fluorescent fusions in the native locus of each gene under the control of its native promoter, followed by the pSG5082 backbone and a truncated copy of the gene. To construct ColFemX-sGFP, we used essentially the same strategy, except that the pFAST3[32] vector was used instead of pSG5082. A *femX* fragment was amplified from COL DNA with the primers femXg_P1 and femXg_P2, digested with KpnI and XhoI, and then cloned into pFAST3 upstream of and in frame with *sgfp* to produce pFAST-femX, which was electroporated into RN4220 and transduced into COL. Strain ColMurG-GFP was obtained by transducing the *murG-gfp* construct from BCBMS001[33] into COL.

The strains ColFemA-mCherry, ColFtsW-mCherry, ColMurJ-mCherry, ColRodA-sGFP, ColsGFP-PBP1 and ColsGFP-PBP3 were constructed by allelic replacement strategies using the pMAD vector. In each case, three DNA fragments (F1, F2 and F3; see Supplementary Table 5) that contained overhangs that were complementary with adjacent fragments were amplified from COL DNA, and then joined by overlap PCR to produce F1–F2–F3 fusion constructs. The full constructs were then amplified by PCR using upstream and downstream primers (P1 and P6, respectively, in each case), digested with the corresponding restriction enzymes and cloned into pMAD. Integration and excision of the pMAD derivatives in COL by a double recombination event that led to allelic exchange was performed as previously described[34]. Supplementary Table 5 describes the cloning strategy for each strain.

To obtain the strains ColFtsW-sGFP and ColMurJ-sGFP, the plasmids pMAD-ftsWsgfp and pMAD-murJsgfp were first constructed. For pMAD-ftsWsgfp, three DNA fragments (F1, F2 and F3) that are each flanked by restriction sites were introduced into pMAD. F1, which contained the 3′ end of *ftsW* minus the stop codon, was amplified from NCTC8325-4 DNA with the primers ftsWg_P1 and ftsWg_P2, and then digested with SmaI and SalI. F2, which contained *sgfp*, was amplified from pTRC99a-P7 with the primers ftsWg_P3 and ftsWg_P4, and then digested with SalI. F3, which contained the region downstream of *ftsW*, was amplified from NCTC8325-4 DNA and digested with SalI and BamHI. Fragments were then sequentially cloned (F1, followed by F3 and finally by F2) into pMAD using the adjacent restriction sites, which produced pMAD-ftsWsgfp. The same strategy was used for pMAD-murJsgfp. F1, which contained the 3′ end of *murJ* minus the stop codon, was amplified from COL DNA using the primers murJg_P1 and murJg_P2, and then digested with SmaI and SalI. F2, which contained *sgfp*, was amplified from pTRC99a-P7 using the primers murJg_P3 and murJg_P4, and then digested with SalI. F3, which contained the last 26 bp of *murJ* and its downstream region, was amplified from COL DNA using the primers murJg_P5 and murJg_P6, and then digested with SalI and BamHI. Fragments were cloned into pMAD, which produced the plasmid pMAD-murJsgfp. The plasmids pMAD-ftsWsgfp and pMAD-murJsgfp were then electroporated to RN4220, transduced to COL and—following allelic replacement—the strains ColFtsW-sGFP and ColMurJ-sGFP were obtained.

The strain ColFtsZ-mCherryi was constructed using the pBCB13 plasmid[35], a derivative of pMAD that allows allelic exchanges in the *spa* locus. In brief, a DNA fragment containing the ribosome binding site, the *ftsZ* gene without its stop codon and a five-amino-acid linker was amplified by PCR from COL with the primers iftsZm_P1 and iftsZm_P2. A second DNA fragment containing *mCherry* was amplified from pBCB4che using the primers iftsZm_P3 and iFtsZm_P4. The two fragments were joined by overlap PCR using the primers iftsZm_P1 and iftsZm_P4, and the resulting construct was then digested with SmaI and XhoI and cloned into pBCB13, downstream of the Pspac promoter, which produced pBCB13-ftsZmch. Strain ColDltC-sGFPi was constructed in a similar fashion: a DNA fragment containing a ribosome binding site, the *dltC* gene without its stop codon and a two-amino-acid linker was amplified by PCR from COL with the primers idltCg_P1 and idltCg_P2. A second DNA fragment containing *sgfp* was amplified from pTRC99a-P7 using the primers idltCg_P3 and idltCg_P4. The two fragments were joined by overlap PCR using the primers idltCg_P1 and idltCg_P4, and the resulting construct was then digested with SmaI and XhoI and cloned into pBCB13 downstream of the Pspac promoter, which produced pBCB13-dltCsgfp. After transduction to COL, plasmid integration and excision at the *spa* locus was performed as previously described[34].

The strain ColFtsZ$^{55–56}$sGFP was constructed using the pCNX replicative plasmid[9] to express an FtsZ–sGFP sandwich fusion. In brief, three DNA fragments (F1, F2 and F3) with overhangs were amplified to construct a *sgfp* fusion inserted within the *ftsZ* coding sequence between codons 55 and 56[36], flanked by 10-amino-acid linkers (Gly-Gly-Gly-Gly-Ser × 2). F1, which contained a ribosome binding site and the first 165 bp of *ftsZ*, was amplified from COL DNA using the primers ftsZswgfp_pCNX_P1 and ftsZswgfp_pCNX_P2. F2, which contained *sgfp* flanked by linker sequences, was amplified from pTRC99a-P7 using the primers ftsZswgfp_pCNX_P3 and ftsZswgfp_pCNX_P4. F3, which contained the remaining 1,008 bp of *ftsZ* (from nucleotide 166 onwards), was amplified from COL DNA using the primers ftsZswgfp_pCNX_P5 and ftsZswgfp_pCNX_P6. The three fragments were joined by overlap PCR, digested with BamHI and EcoRI, and cloned into pCNX to give pCN-ftsZ$^{55–56}$sGFP, which was then transduced into COL. The resulting strain was named ColFtsZ$^{55–56}$sGFP.

The strain ColMraY-sGFP was constructed using the pCN51 replicative plasmid to express an MraY–sGFP sandwich fusion. In brief, three DNA fragments (F1, F2 and F3) with overhangs were amplified to construct a fusion with *sgfp* inserted within the *mraY* coding sequence, between codons 220 and 221. F1, which contained a ribosome binding site and the first 660 bp of *mraY,* was amplified from COL DNA using the primers mraYg_P1 and mraYg_P2. F2, which contained *sgfp* minus the stop codon, was amplified from pTRC99a-P7 using the primers mraYg_P3 and mraYg_P4. F3, which contained the last 306 bp of *mraY,* was amplified from COL using the primers mraYg_P5 and mraYg_P6. The three fragments were joined by overlap PCR and digested with SmaI and cloned into pCN51, which produced pCN-mraYsgfp.

The strains ColWZ and ColJZ were constructed by transducing the plasmids pMAD-ftsWmch and pMAD-murJmch, respectively, into BCBAJ020. ColP1Z was constructed by transducing pBCB13-ftsZmch into ColsGFP-PBP1. The strains ColWgZm and ColJgZm were constructed by transducing the plasmids pMAD-ftsWsgfp and pMAD-murJsgfp, respectively, into ColFtsZ-mCherryi. The strain ColWJ was obtained by transducing pMAD-murJsgfp to ColFtsW-mCherry. In each case, allelic replacement was performed as described earlier.

To construct ColZZ, an *ftsZ–mCherry* fusion was amplified from the genomic DNA of ColFtsZ-mCherryi with the primers ftsZm_pCNX_P1 and ftsZm_pCNX_P2, digested with BamHI and EcoRI, and then cloned in pCNX downstream of the Pcad promoter, giving plasmid pCN-ftsZmch. This plasmid was then electroporated into RN4220 and transduced to BCBAJ020, which produced strain ColZZ.

To study co-localization between DivIB and FtsZ or MurJ, an *yfp–divIB* fusion was constructed and cloned into pCNX. In brief, a DNA fragment that contained *yfp* minus the stop codon and a 3′ terminal overhang was amplified from pMUTINYFPKan[37] with the primers ydivIB_pCNX_P1 and ydivIB_pCNX_P2. A second DNA fragment that contained the full *divIB* gene with a 5′ overhang was amplified from COL DNA with the primers ydivIB_pCNX_P3 and ydivIB_pCNX_P4. The two fragments were then joined by overlap PCR, digested with SmaI and KpnI, and cloned into pCNX downstream of Pcad, which produced the plasmid pCN-yfpDivIB. This plasmid was transduced to BCBAJ020 and ColMurJ-mCherry to produce the strains ColZIB and ColJIB, respectively.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Construction of *S. aureus* strains containing antisense RNA vectors. To construct strains carrying antisense RNA vectors, 250-bp fragments of *divIB* or *divIC* genes were amplified from COL DNA with the primer pairs ASdivIB_P1 and ASdivIB_P2, and ASdivIC_P1 and AS_DivIC_P2, respectively, then digested with EcoRI and BamHI and cloned in an antisense direction—relative to the xylose-inducible T5X promoter—into pEPSA5 to produce pAS-DivIB and pAS-DivIC. These plasmids were then transduced into ColJZ to produce ColJZpAS-DivIB and ColJZpAS-DivIC, respectively. Additionally, phage lysates were obtained from the AS-022 and AS-185 strains[21] carrying the antisense RNA pEPSA vectors pAS-022 and pAS-185, which target *ftsA* and *ftsL*, respectively. pAS-022 was transduced to ColWZ and ColP1Z to produce the strains ColWZpAS-FtsA and ColP1ZpAS-FtsA, respectively. pAS-185 was transduced to ColJZ, ColWZ, ColP1Z and ColWJ to produce the strains ColJZpAS-FtsL, ColWZpAS-FtsL, ColP1ZpAS-FtsL and ColWJpAS-FtsL, respectively. Control strains for these experiments were obtained by transducing the empty vector pEPSA5 into ColJZ, ColWZ, ColP1Z and ColWJ to produce the strains ColJZpEPSA, ColWZpEPSA, ColP1ZpEPSA and ColWJpEPSA, respectively.

Growth curves of *S. aureus* strains. Overnight cultures of COL strains that encoded fluorescent derivatives of PG synthesis enzymes were back-diluted to $OD_{600nm} = 0.02$ in TSB and grown at 37 °C for 11 h with OD_{600nm} measurements taken every hour. Doubling times were calculated for each strain during its exponential growth phase.

Minimum inhibitory concentration assays. Minimum inhibitory concentrations of relevant antimicrobial compounds were determined by broth microdilution in sterile 96-well plates. The medium used was TSB that contained a series of twofold dilutions of each compound. Cultures of *S. aureus* strains and mutants were added at a final density of approximately 5×10^5 CFU ml^{-1} to each well. Wells were reserved in each plate for sterility control (no cells added) and cell viability (no compound added). Plates were aerobically incubated at 37 °C. Endpoints were assessed visually after 24 and 48 h. All assays were done in triplicate.

Western blotting. The *S. aureus* strains ColJZpEPSA, ColJZpAS-DivIB and ColJZpAS-DivIC were grown overnight, back-diluted to 1:200 in fresh TSB and incubated at 37 °C until they reached an OD_{600nm} of approximately 0.2. At this point, xylose was added to the medium at 4% to enable the expression of the antisense RNA fragments. After 1 h of antisense expression, cells were collected and broken with glass beads in a FastPrep FP120 cell disrupter (Thermo Electro). Samples were centrifuged to remove unbroken cells and debris, and the total protein content of the clarified lysates was determined using the Bradford method and bovine serum albumin as a standard (BCA Protein Assay Kit, Pierce). Equal amounts of total protein from each sample were separated on 10% SDS–PAGE at 80 V and then transferred to Hybond-P PVDF membrane (GE Healthcare) using a semi-dry transfer cell (Biorad), according to standard western blotting techniques. Membranes were cut to separate the PBP2A region from the DivIB or DivIC region. DivIB and DivIC proteins were detected using specific polyclonal antibodies at 1:5000 and 1:10000 dilutions, respectively. PBP2A was detected using the antibody from a Slidex MRSA detection kit (Biomerieux) at 1:500 dilution. Protein bands were visualized using the ECL Prime Detection Reagents (GE Healthcare), according to the manufacturer's instructions.

***S. aureus* imaging by fluorescence microscopy.** Super-resolution SIM imaging was performed using an Elyra PS.1 microscope (Zeiss) with a Plan-Apochromat 63×/1.4 oil DIC M27 objective. SIM images were acquired using five grid rotations, unless stated otherwise, with a 34-μm grating period for the 561-nm laser (100 mW), a 28-μm period for the 488-nm laser (100 mW) and a 23-μm period for the 405-nm laser (50 mW). Images were captured using a Pco.edge 5.5 camera and reconstructed using ZEN software (black edition, 2012; version 8.1.0.484) on the basis of a structured illumination algorithm, with synthetic, channel specific optical transfer functions and noise filter settings ranging from −6 to −8.

Wide-field fluorescence microscopy was performed using a Zeiss Axio Observer microscope with a Plan-Apochromat 100×/1.4 oil Ph3 objective. Images were acquired with a Retiga R1 CCD camera (QImaging) using Metamorph 7.5 software (Molecular Devices).

Unless stated otherwise, for fluorescence microscopy experiments overnight cultures of *S. aureus* strains were back-diluted to 1:200 in fresh medium with appropriate inducers and allowed to grow until they reached an OD_{600nm} of approximately 0.6 before being collected and washed with phosphate buffered saline (PBS). Cells were then placed on microscope slides covered with a thin layer of agarose (1% in PBS) and imaged by SIM or wide-field microscopy.

To assess whether MurJ localization was dependent on interaction with its substrate, the strain ColMurJ-mCherry was grown until it reached an OD_{600nm} of 0.4 and incubated with 3-{1-[(2,3-dimethylphenyl)methyl]piperidin-4-yl}-1-methyl-2-pyridin-4-yl-1*H*-indole (DMPI, a gift from Merck) at 3 μg ml^{-1} for 30 min, and then imaged by wide-field fluorescence microscopy as described earlier.

For antisense RNA experiments, strains were grown until they reached an OD_{600nm} of 0.1–0.2, at which point expression of the antisense RNA fragments was induced with xylose (Apollo Scientific) at a final concentration of 4% for 1 h. Cells were then collected and washed with PBS to remove xylose, mounted on microscope slides as described earlier and imaged by wide-field fluorescence microscopy. Assays were done in triplicate.

To evaluate localization of peptidoglycan synthesis activity, *S. aureus* cells were given a pulse of HADA[17] (250 μM) for 1 min and then washed two times with PBS. Cells were then placed on an agarose pad and visualized by wide-field microscopy. Assays were done in triplicate.

To label *S. aureus* membranes, cells were incubated with Nile Red (Invitrogen) at a final concentration of 10 μg ml^{-1} for 5 min at room temperature, washed with PBS and then mounted on microscope slides.

For time-lapse experiments, all cultures were grown overnight in TSB. For ColFtsZ$^{55–56}$sGFP and ColpSGEzrA-GFP, the medium was supplemented with kanamycin (150 μg ml^{-1}) and erythromycin (10 μg ml^{-1}), respectively. Cultures grown overnight were diluted to 1:200 in fresh TSB medium, without antibiotics but supplemented with appropriate inducers ($CdCl_2$ at 0.1 μM for ColFtsZ$^{55–56}$sGFP; IPTG at 0.5 mM for ColWgZm and ColJgZm). After being collected, cells were re-suspended in M9 microscopy medium (KH_2PO_4 3.4 g l^{-1}, VWR; K_2HPO_4 2.9 g l^{-1}, VWR; di-ammonium citrate 0.7 g l^{-1}, Sigma-Aldrich; sodium acetate 0.26 g l^{-1}, Merck; glucose 1%, Merck; $MgSO_4$ 0.7 mg l^{-1}, Sigma-Aldrich; $CaCl_2$ 7 mg l^{-1}, Sigma-Aldrich; casamino acids 1%, Difco; 1× MEM amino acids, Thermo Fisher Scientific; MEM vitamins 1×, Thermo Fisher Scientific). The medium was supplemented when required with IPTG 0.5 mM, $CdCl_2$ 0.1 μM, DMPI 8 μg ml^{-1} or PC190723 5 μg ml^{-1}. Cultures were then spotted on a pad of 1.5% agarose in the same supplemented medium, and mounted in a gene frame on a microscope slide. The time between the cells contacting the pad and the start of acquisition was 5 min in all conditions.

HADA incorporation microscopy assay. To assess whether MurJ activity was required for HADA incorporation, the strains COL and ColDltC-GFPi, which expresses a cytoplasmic DltC–sGFP fusion and can therefore be easily distinguished from COL under the microscope, were grown to an OD_{600nm} of 0.4. At this point, each culture was separated into two flasks and either DMPI (3 μg ml^{-1} in DMSO) or DMSO (0.2% final concentration) was added to the cultures. After 25 min of incubation, HADA (500 μM) was added to each culture for 5 min. Cells were then collected, washed twice with PBS (supplemented with DMPI when applicable) and DMSO-treated COL cells were mixed with DMPI-treated ColDltC-sGFPi cells. To exclude the possibility that the expression of DltC–sGFP affected the results, a reverse experiment was performed in which cells of DMSO-treated ColDltC-sGFPi were mixed with DMPI-treated COL. These mixtures of two strains were then imaged by fluorescence microscopy as described earlier.

Functionality test for EzrA–GFP construct. The strains COL and COLΔEzrA were grown overnight in TSB, and the strain ColpSGEzrA-GFP was grown in TSB plus erythromycin 10 μg ml^{-1}, at 37 °C with aeration. Cultures grown overnight were diluted to 1:200 in fresh TSB and incubated at 37 °C with aeration. Once the OD_{600nm} reached approximately 0.6, cells were pelleted, resuspended in PBS and spotted on a thin layer of 1.2% agarose in PBS. Cultures were imaged by phase contrast, then the single cells were identified and cell area measured using eHooke (see below; a lack of EzrA produces larger cells[38]).

Imaging of FtsZ$^{55–56}$sGFP treadmilling by SIM. To visualize FtsZ$^{55–56}$sGFP movement in the strain ColFtsZ$^{55–56}$sGFP, and EzrA–GFP movement in the strain ColpSGEzrA-GFP, 50 frames of SIM images were acquired with 5-s intervals (total time of acquisition of 250 s) using the 488-nm laser at 50% and a 50-ms exposure time. For FtsZ, after this acquisition, 3 time-lapse images were taken 15 min apart to check whether the corresponding rings were constricting. These experiments were performed with or without PC190723 at 5 μg ml^{-1}. To analyse the movement of FtsZ or EzrA, the 50 frames were aligned and 1-pixel lines were drawn over the rings and plotted in a kymograph.

Visualization of divisome ring constriction by SIM. To follow FtsZ$^{55–56}$sGFP or EzrA–GFP ring constriction over time, ColFtsZ$^{55–56}$sGFP or ColpSGEzrA-GFP dividing cells were imaged by SIM every 5 min for 1 h using the 488-nm laser at 50% power and a 50-ms exposure time. These experiments were performed with or without PC190723 (5 μg ml^{-1}; laser power 50%), and with DMPI (8 μg ml^{-1}; laser power 100%) for FtsZ only. After SIM image reconstruction, cropped stacks containing individual cells were used to plot Z-ring constriction kymographs over the 1-h period by using Fiji software[39] to draw a 3-pixel line, which crossed the fluorescent signal, across the diameter of the cell. The percentage of cells in the presence of PC190723 for which the FtsZ ring did not constrict, constricted with defects or constricted in relation to their initial ring size was plotted using GraphPad Prism 6 (https://www.graphpad.com/scientific-software/prism/).

To follow the constriction of FtsW–sGFP and MurJ–sGFP rings in the strains ColFtsW-sGFP and ColMurJ-sGFP, respectively, cultures were spotted on an

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

M9 pad with or without PC190723, and then imaged by SIM. The 488-nm laser was used at 100% intensity, with an exposure time of 50-ms. To decrease photodamage caused by high laser power, images were acquired every 10 min, instead of 5 every min, for 1 h. These settings were also applied to ColFtsZ$^{55-56}$sGFP for comparison. For MurJ–sGFP time-lapse images were also taken at 5-min intervals to confirm the absence of biphasic behaviour of the constricting ring. Only cells that did not show a MurJ–sGFP signal in the first frame but which did present this signal in the second frame were analysed, to ensure that the entire constriction process was observed (that is, to confirm that the absence of a biphasic behaviour was not the result of only imaging cells in the later stages of cell division). After SIM image reconstruction, cropped stacks containing individual cells were used to plot constriction kymographs as described earlier.

Time-lapse stacks of ColFtsZ$^{55-56}$sGFP, ColFtsW-sGFP and ColMurJ-sGFP in the presence of PC190723 were used to count the number of cells that had fusion at the midcell in the first frame and to determine the percentage of those cells that constricted, constricted with defects or did not constrict in the presence of PC190723 for 60 min. Data were plotted using GraphPad Prism 6.

To observe the constriction of FtsW–sGFP and MurJ–sGFP rings in cells expressing FtsZ–mCherry, the strains ColWgZm and ColJgZm in M9 medium supplemented with IPTG were imaged by SIM every 10 min for 1 h, using the 488-nm laser and the 561-nm laser (100% laser power, 50-ms and 50% laser power, 50-ms, respectively). Cells of ColWgZm in which FtsZ–mCherry showed biphasic constriction behaviour and cells of ColJgZm in which a MurJ–sGFP signal appeared on the third frame were used to measure the ring diameter for each protein in all frames.

Inhibition of the cell cycle assays. A culture of strain COL grown overnight was back-diluted to 1:200 and grown until it reached an OD_{600nm} of approximately 0.4, at which point either DMPI[12] (3 μg ml^{-1}), PC190723 (2.5 μg ml^{-1}), oxacillin (1,000 μg ml^{-1}, Sigma-Aldrich) or DMSO (0.2%) was added to the medium. Cells were then grown for 30 min with each inhibitor, collected and labelled with Nile Red, as described earlier, before being imaged by SIM. Cells were sorted into each phase of the cell cycle (phase 1, phase 2 or phase 3), as previously described[9]. Assays were performed in triplicate.

Automatic cell-imaging analysis. Analysis of phase-contrast and fluorescence images was performed using eHooke software that we developed in-house. For cell segmentation, eHooke uses phase-contrast images and applies the isodata algorithm[40] for automatic thresholding to find a pixel intensity value that separates the pixels that correspond to the background from those pixels that correspond to cells. Using this threshold, the software then creates a binary mask that is used to compute the Euclidean distances[41] of each pixel in order to find the centres of individual regions inside the mask. The software then expands those centres to define individual cell regions using the watershed algorithm[42].

To measure cell areas and volumes, eHooke first defines each individual cell region, computes the area by counting the number of pixels inside the region and the volume by measuring the long and short axes of the cell. Cell volume is then derived assuming a prolate spheroid shape as previously described[9].

To calculate fluorescent ratios between septal and membrane signals, eHooke was used to define the different regions of the cell (the membrane, cytoplasm and septum) in images obtained by wide-field fluorescence microscopy. The membrane is defined by dilating the outline of the cell towards its inside. To separate the septum from the cytoplasm, the software uses the isodata algorithm[40] to find the brightest region inside the cell. This region is then defined as the septum. To measure the median fluorescence, only the 25% brightest pixels of the septum were considered, to remove possible misidentified pixels from the measurement. Only cells with a closed septum were selected for measurements. Fluorescence ratio values were then calculated according to the equation:

$$\text{Fluorescence ratio} = \frac{\text{median(septum)} - \text{background}}{\text{median(membrane)} - \text{background}}$$

To measure the PCC between two fluorescent proteins in a strain, images of each fluorescence channel were aligned and loaded side-by-side in eHooke. After automatic cell segmentation, cells that showed an FtsZ signal at the septum were selected for PCC measurements. The pixels corresponding to each cell were isolated and PCC values between channels were then calculated using an equation adapted from ref. 43:

$$\text{PCC} = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_i (X_i - \overline{X})^2}\sqrt{\sum_i (Y_i - \overline{Y})^2}}$$

in which X_i and Y_i correspond to each pixel intensity for two fluorescence channels and $\overline{X}$ and $\overline{Y}$ correspond to the mean intensities of those channels.

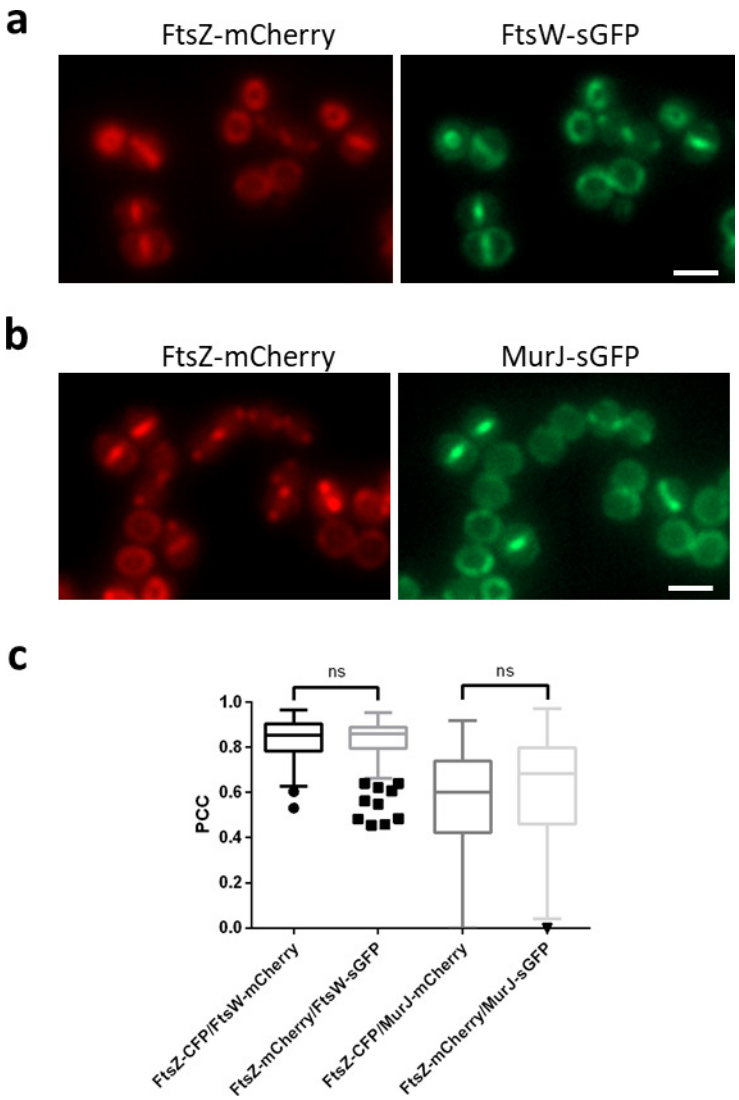
Statistical analysis. Statistical analyses were done using GraphPad Prism 6 (GraphPad Software). Unpaired Student's *t*-tests were used to evaluate the differences of cellular volumes, and to compare fluorescence ratios between peripheral and septal wall signal intensities. Mann–Whitney *U* tests were used to compare differences between PCC non-normal distributions obtained in co-localization experiments. P values ≤ 0.05 were considered as significant for all analysis performed and are indicated with asterisks: $^{*}P \leq 0.05$, $^{**}P \leq 0.01$ and $^{***}P \leq 0.001$.

Code availability. Code is available from: https://github.com/BacterialCellBiologyLab/Bruno-Saraiva-2017.

Data availability. Source data for Figs 1–3 and 5 and Extended Data Figs 1–4, 7 and 9 are provided with the paper. All other data are available from the corresponding author upon reasonable request.

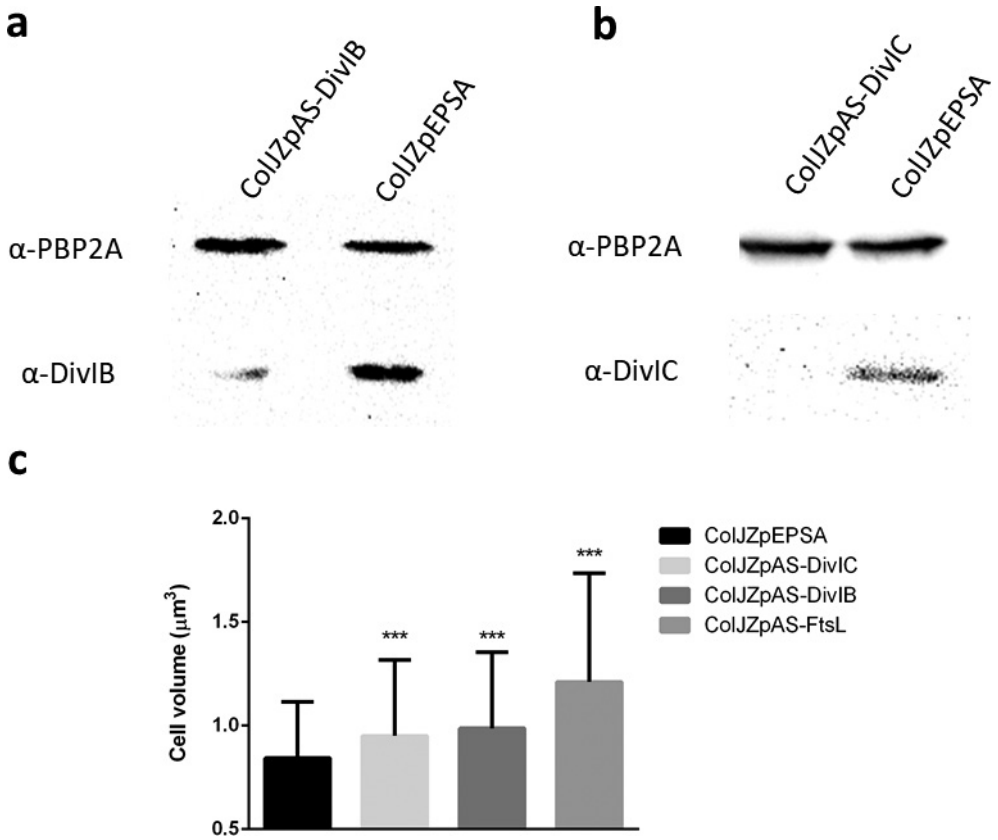
29. Kraemer, G. R. & Iandolo, J. J. High-frequency transformation of *Staphylococcus aureus* by electroporation. *Curr. Microbiol.* **21,** 373–376 (1990).
30. Oshida, T. & Tomasz, A. Isolation and characterization of a Tn*551*-autolysis mutant of *Staphylococcus aureus*. *J. Bacteriol.* **174,** 4952–4959 (1992).
31. Pinho, M. G. & Errington, J. A *divIVA* null mutant of *Staphylococcus aureus* undergoes normal cell division. *FEMS Microbiol. Lett.* **240,** 145–149 (2004).
32. Nair, D. R. *et al.* Characterization of a novel small molecule that potentiates β-lactam activity against gram-positive and gram-negative pathogens. *Antimicrob. Agents Chemother.* **59,** 1876–1885 (2015).
33. Mann, P. A. *et al.* Murgocil is a highly bioactive staphylococcal-specific inhibitor of the peptidoglycan glycosyltransferase enzyme MurG. *ACS Chem. Biol.* **8,** 2442–2451 (2013).
34. Arnaud, M., Chastanet, A. & Débarbouillé, M. New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. *Appl. Environ. Microbiol.* **70,** 6887–6891 (2004).
35. Pereira, P. M., Veiga, H., Jorge, A. M. & Pinho, M. G. Fluorescent reporters for studies of cellular localization of proteins in *Staphylococcus aureus*. *Appl. Environ. Microbiol.* **76,** 4346–4353 (2010).
36. Moore, D. A., Whatley, Z. N., Joshi, C. P., Osawa, M. & Erickson, H. P. Probing for binding regions of the FtsZ protein surface through site-directed insertions: discovery of fully functional FtsZ-fluorescent proteins. *J. Bacteriol.* **199,** e00553–16 (2016). 10.1128/jb.00553-16
37. Atilano, M. L. *et al.* Teichoic acids are temporal and spatial regulators of peptidoglycan cross-linking in *Staphylococcus aureus*. *Proc. Natl Acad. Sci. USA* **107,** 18991–18996 (2010).
38. Jorge, A. M., Hoiczyk, E., Gomes, J. P. & Pinho, M. G. EzrA contributes to the regulation of cell size in Staphylococcus aureus. *PLoS ONE* **6,** e27542 (2011).
39. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).
40. Velasco, F. R. D. Thresholding using the ISODATA clustering algorithm. *IEEE Trans. Syst. Man Cybern.* **10,** 771–774 (1980).
41. Falcão, A. X., Stolfi, J. & de Alencar Lotufo, R. The image foresting transform: theory, algorithms, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **26,** 19–29 (2004).
42. Roerdink, J. B. T. M. & Meijster, A. The watershed transform: definitions, algorithms and parallelization strategies. *Fundam. Inform.* **41,** 187–228 (2000).
43. Dunn, K. W., Kamocka, M. M. & McDonald, J. H. A practical guide to evaluating colocalization in biological microscopy. *Am. J. Physiol. Cell Physiol.* **300,** C723–C742 (2011).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



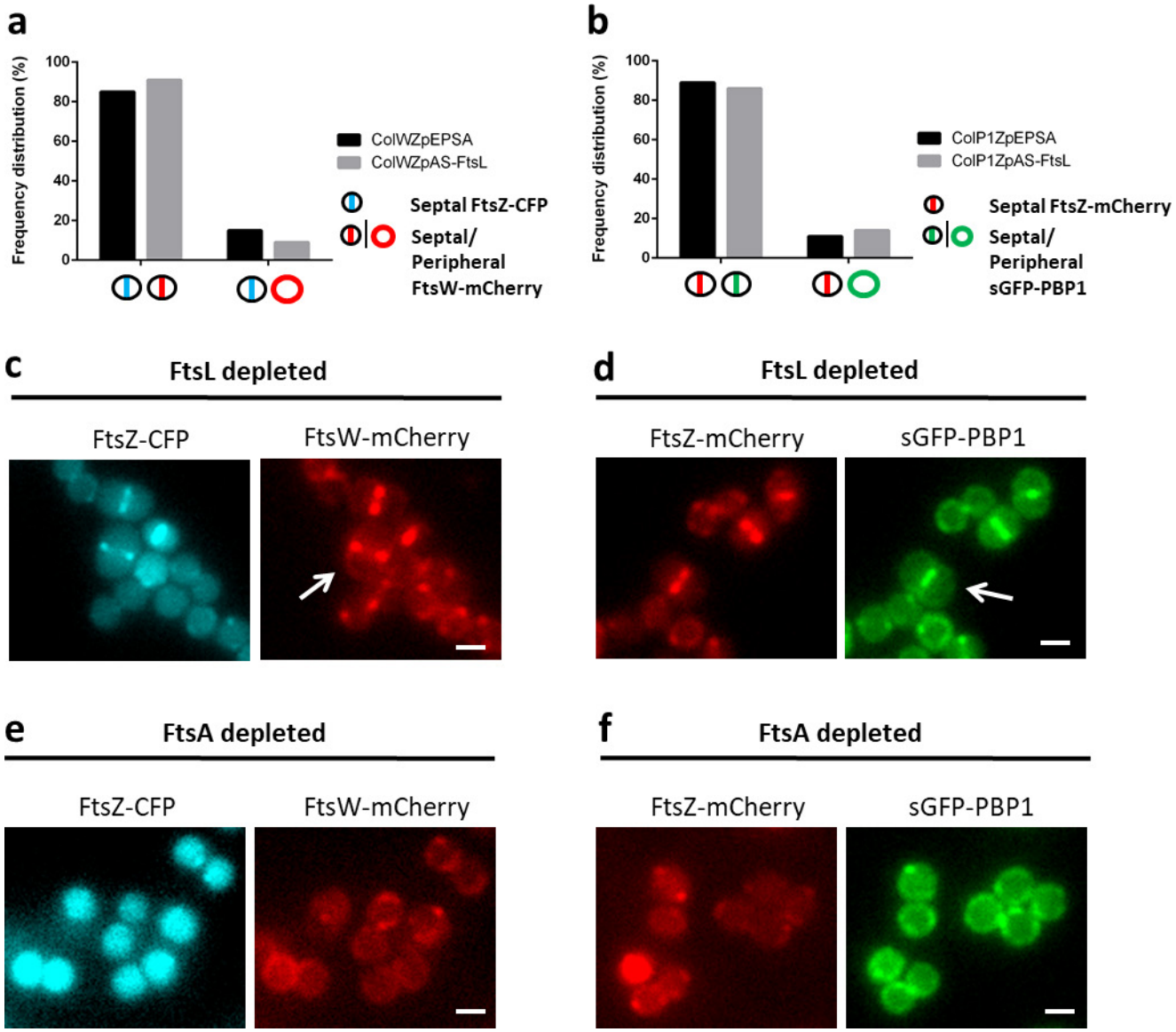
Extended Data Figure 1 | Switching fluorescent tags has no effect on protein co-localization data. a, **b**, COL strains that express FtsZ–mCherry and FtsW–sGFP (**a**) or FtsZ–mCherry and MurJ–sGFP (**b**) were compared to strains that express FtsZ–CFP and FtsW–mCherry, and FtsZ–CFP and MurJ–mCherry, respectively (described in Fig. 2b, c). Scale bars, 2 μm. **c**, PCC values between fluorescence channels for each protein fusion pair were calculated for cells that showed septal FtsZ localization. From left to right, $n = 138$, 136, 133 and 139 cells. Negative PCC values are represented as 0. Data are represented as box-and-whisker plots in which boxes correspond to the first-to-third quartiles, lines inside the boxes indicate the median, and the ends of whiskers and outliers follow a Tukey representation. Statistical analysis was performed using a two-sided Mann–Whitney *U* test; ns, not significant. Images in **a** and **b** are representative of three biological replicates.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



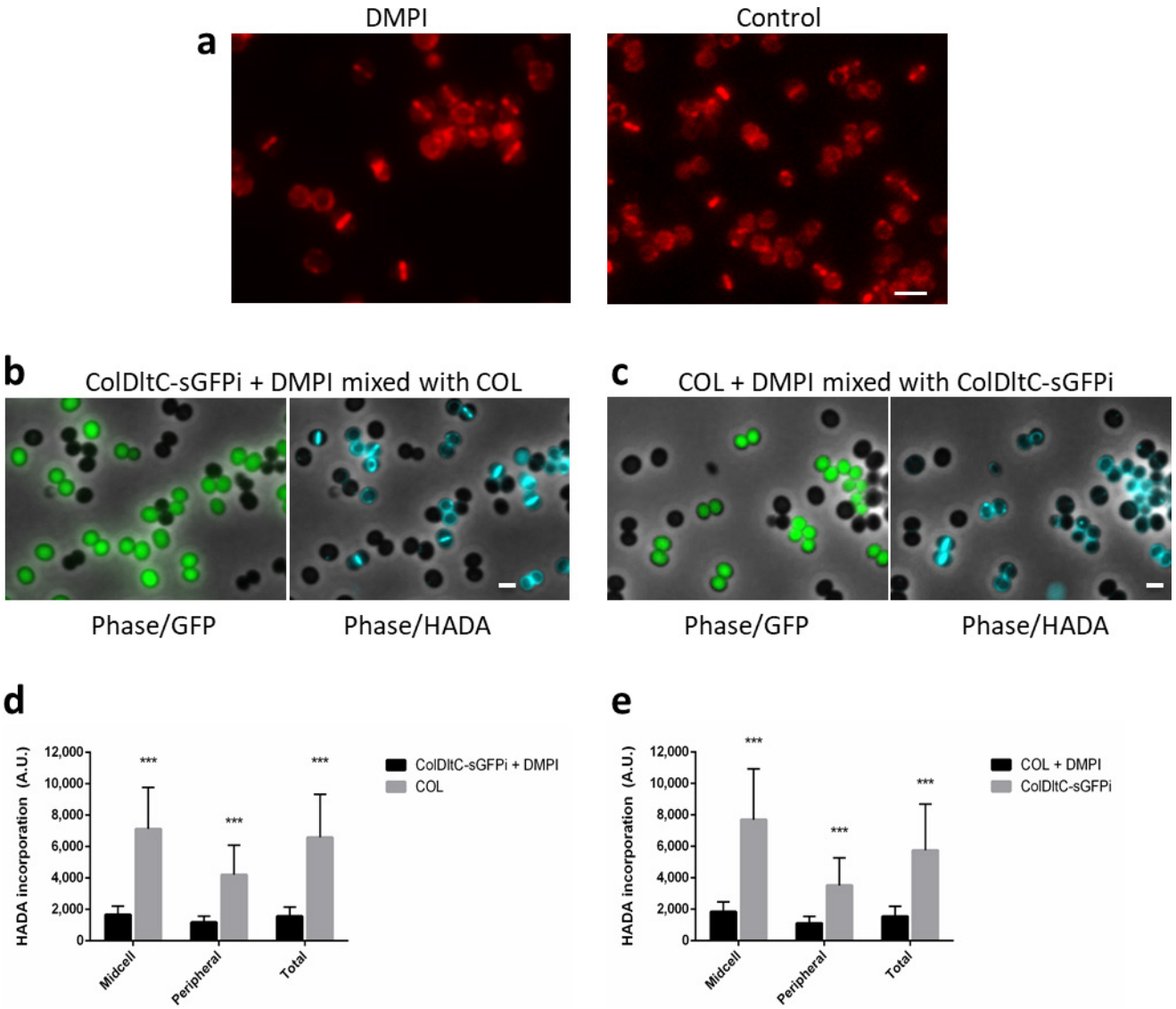
Extended Data Figure 2 | Antisense RNA fragments that target the DivIB–DivIC–FtsL complex increased cell volume and decreased protein expression. a, Western blot that shows total protein extracts of ColJZpAS-DivIB after 1 h of antisense induction, and control ColJZpEPSA probed with antibodies against either PBP2A (loading control; upper bands) or DivIB (lower bands). **b**, Western blot that shows total protein extracts of ColJZpAS-DivIC after 1 h of antisense induction, and control ColJZpEPSA probed with antibodies against either PBP2A (loading control; upper bands) or DivIC (lower bands). Images in **a** and **b** are representative of three biological replicates. For gel source data, see Supplementary Fig. 1. **c**, Cell volume of cells that expressed antisense RNA against *ftsL*, *divIB* or *divIC*, or carried vector pEPSA (left to right, $n = 421$, 379, 279 and 361 cells). Data represented in column graphs in which the height of the column represents the mean and the whiskers are the s.d. Statistical analysis was performed using a two-sided unpaired *t*-test. $^{***}P < 0.001$; first versus second column, $P = 3.50 \times 10^{-6}$; first versus third column, $P = 3.80 \times 10^{-8}$; first versus fourth column, $P = 1.27 \times 10^{-29}$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



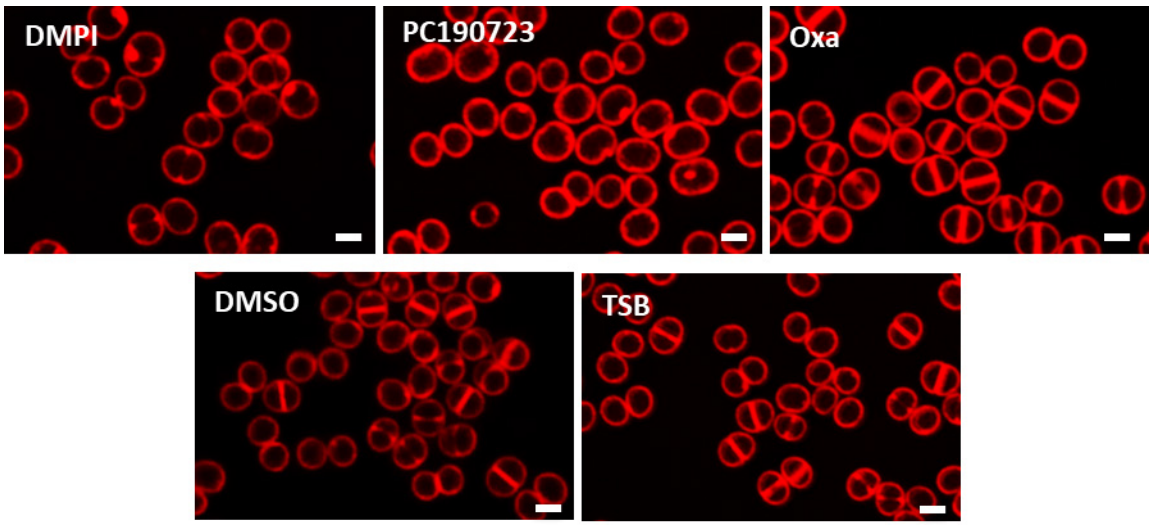
Extended Data Figure 3 | FtsW and PBP1 recruitment to the divisome is independent of DivIB–DivIC–FtsL complexes. The strain ColWZpAS-FtsL, which harboured the FtsZ–CFP and FtsW–mCherry fluorescent fusion pair, and the strain ColP1ZpAS-FtsL, which harboured the FtsZ–mCherry and sGFP–PBP1 fluorescent fusion pair, were depleted of FtsL expression using antisense RNA and imaged by wide field fluorescence microscopy. **a**, Frequency of FtsZ–CFP and FtsW–mCherry co-occurrence in ColWZpAS-FtsL as compared to control ColWZpEPSA ($n = 200$ for each). **b**, Frequency of FtsZ–mCherry and sGFP–PBP1 co-occurrence in ColP1ZpAS-FtsL and in control ColP1ZpEPSA ($n = 200$ for each). **c**, **d**, Very large FtsL-depleted cells were observed either where FtsW–mCherry (**c**) or where sGFP–PBP1 (**d**) co-localized with the FtsZ fusion at the septum (arrows). **e**, **f**, Inhibition of divisome assembly at an early stage by depletion of FtsA expression in either ColWZpAS-FtsA or ColP1ZpAS-FtsA prevented the recruitment of FtsW–mCherry **(e)** or sGFP–PBP1 **(f)** to the septum, concomitant with FtsZ destabilization. Images in **c**–**f** are representative of three biological replicates. Scale bars, 1 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



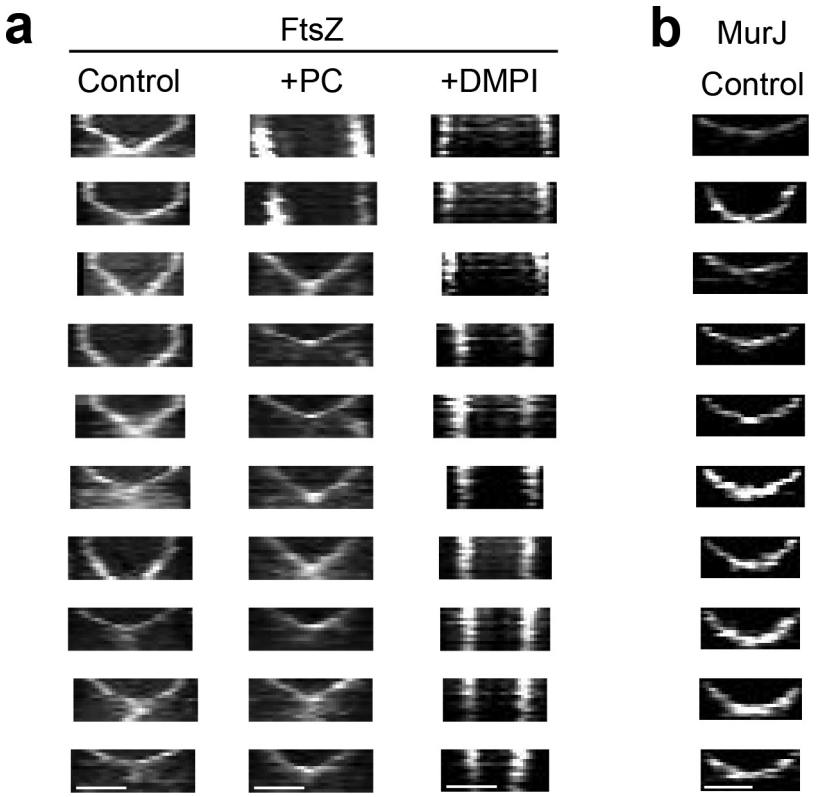
Extended Data Figure 4 | Inhibition of MurJ activity does not prevent its recruitment to the midcell but impairs PG synthesis. a, Fluorescence microscopy images of ColMurJ-mCherry cells grown in the presence (left) or absence (right) of the MurJ inhibitor DMPI for 30 min at 2× minimum inhibitory concentration. Scale bar, 2 μm. **b**, **c**, Fluorescence microscopy images showing mixed cultures of either DMPI-treated ColDltC-sGFPi cells mixed with COL cells (**b**) or DMPI-treated COL cells mixed with ColDltC-sGFPi cells (**c**), after incubation with HADA. The two cultures, which can be easily distinguished owing to GFP expression in ColDltC-sGFPi, were mixed on the same slide to decrease the fluorescence variation of HADA signal that could result from imaging conditions. Data show that HADA incorporation (that is, PG synthesis) is greatly reduced in the presence of DMPI. Phase-contrast–GFP channel overlays are shown on the left, and phase-contrast–HADA channel overlays are shown on the right. Scale bars, 1 μm. **d**, **e**, HADA fluorescence signal measured at the midcell (midcell), the periphery (peripheral) or over the entire cell (total) of DMPI-treated ColDltC-sGFPi cells mixed with COL cells (**d**) or DMPI-treated COL cells mixed with ColDltC-sGFPi cells (**e**). Images in **a**–**c** are representative of three biological replicates. Data in **d** and **e** are represented as column plots ($n = 100$ cells for each column) in which the height of the column is the mean and the whiskers indicate s.d. Statistical analysis was performed using two-sided unpaired *t*-tests. $***P < 0.001$. **d**, $P = 2.34 \times 10^{-38}$ for the midcell; $P = 1.81 \times 10^{-29}$ for the periphery; $P = 9.22 \times 10^{-34}$ for the entire cell. **e**, $P = 1.74 \times 10^{-33}$ for the midcell; $P = 8.77 \times 10^{-25}$ for the periphery; $P = 7.60 \times 10^{-26}$ for the entire cell.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



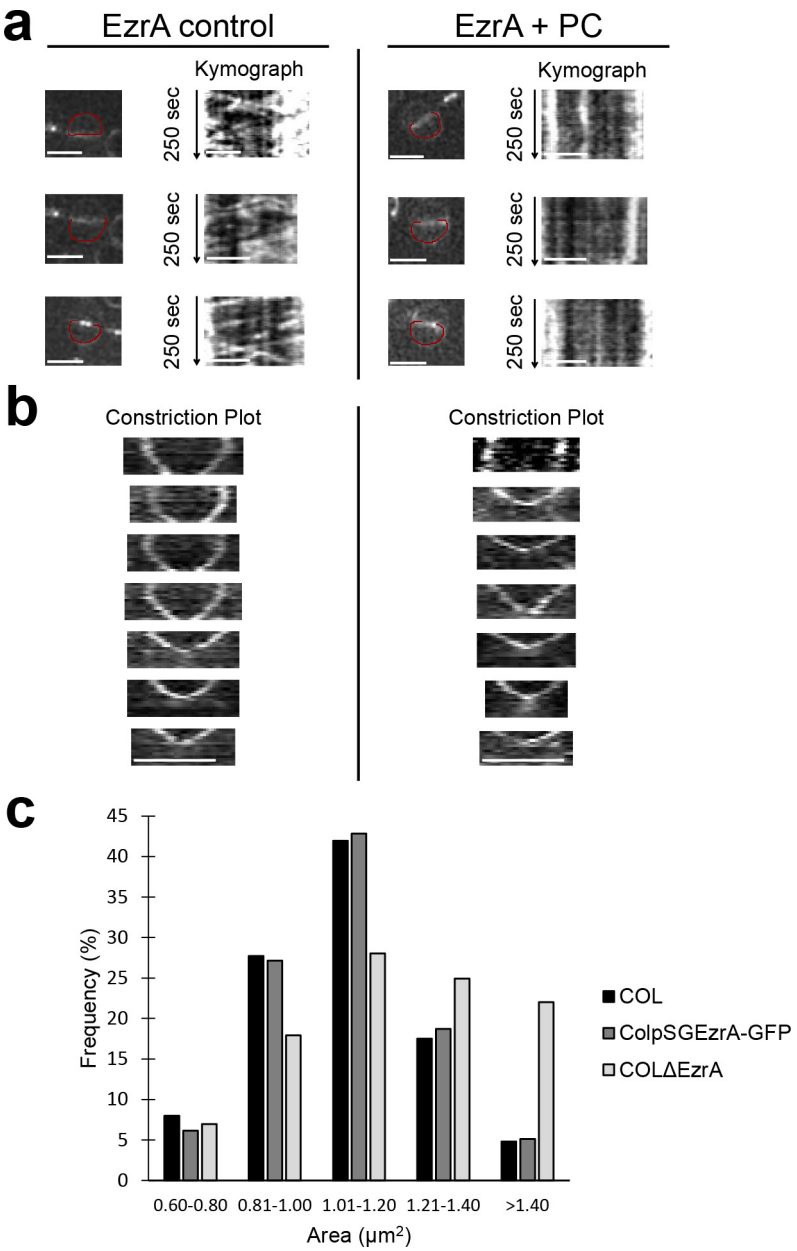
Extended Data Figure 5 | Effect of antibiotics on cell-cycle progression of *S. aureus*. SIM images of Nile-Red-stained COL cells treated with DMPI, PC190723, oxacillin (Oxa), DMSO or TSB (mock-treated controls) for the duration of one cell cycle (30 min). Images are representative of three biological replicates. Scale bars, 1 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



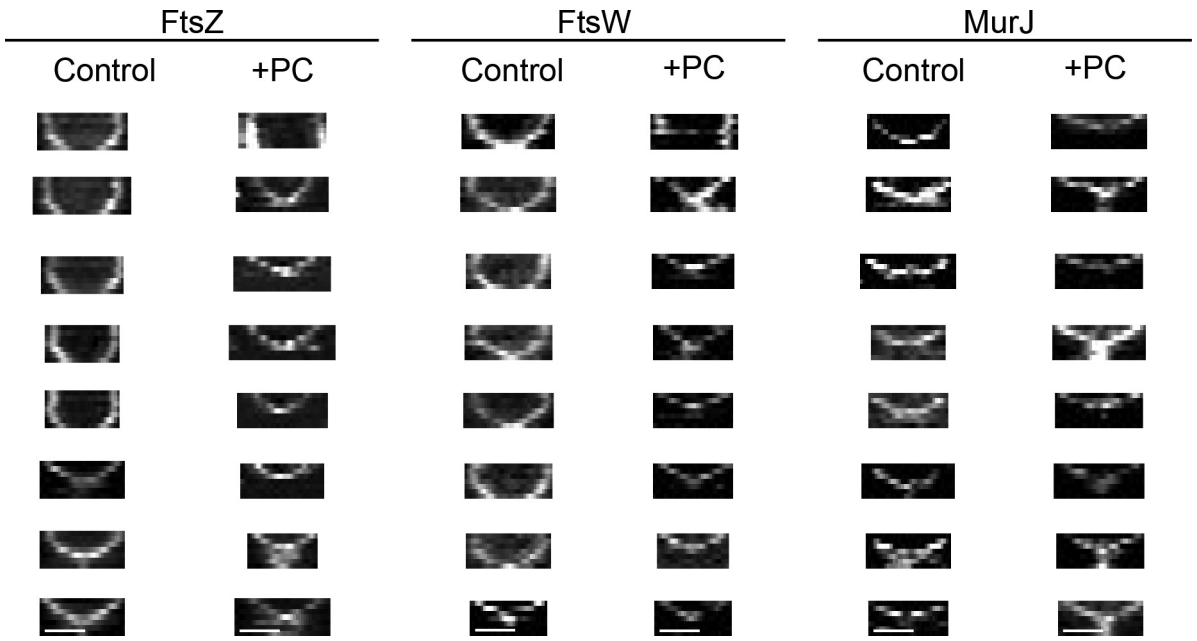
Extended Data Figure 6 | Kymographs that show the constriction of FtsZ$^{55–56}$sGFP and MurJ–sGFP rings. a, Kymographs of 10 cells per column that show the constriction of FtsZ$^{55–56}$sGFP rings, obtained by imaging ColFtsZ$^{55–56}$sGFP cells every 5 min for a total of 60 min (laser power 50%), in the absence (control) or presence of either PC190723 (+PC) or DMPI (+DMPI). Because *S. aureus* cells are not synchronised, cells at all stages of cytokinesis can be observed. Larger FtsZ$^{55–56}$sGFP rings had a biphasic behaviour (no or slow constriction, followed by fast constriction) whereas smaller rings that are further ahead in the cell cycle were only observed undergoing the fast constriction step. The addition of PC190723 inhibited the constriction of larger rings (top two kymographs) but not of smaller rings, which were able to complete cytokinesis. The addition of DMPI completely blocked constriction of FtsZ$^{55–56}$sGFP rings of all sizes. **b**, Kymographs showing constriction of MurJ–sGFP rings in ColMurJ-sGFP cells imaged every 5 min for a total of 60 min (laser power 100%). Cells in which a MurJ–sGFP signal appeared on the second frame were chosen for analysis to ensure that the entire constriction process was imaged. Fast constriction started immediately on the arrival of MurJ–sGFP at the division septum and therefore rings did not show biphasic behaviour. Data are representative of three biological replicates. Scale bars, 0.5 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



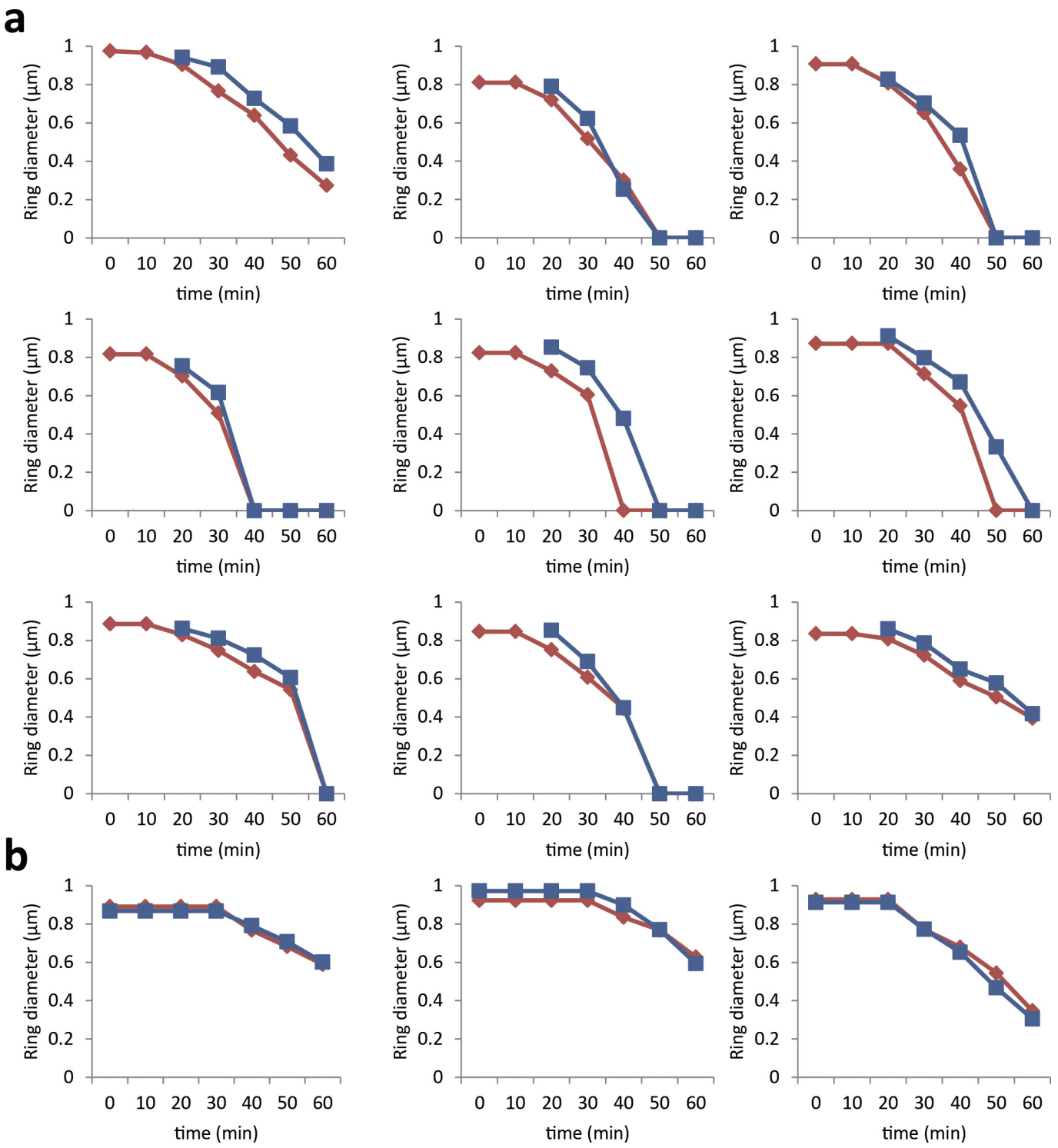
Extended Data Figure 7 | The Z ring protein EzrA shows impaired treadmilling in the presence of PC190723 and biphasic ring constriction. a, ColpSGEzrA-GFP cells that express a functional EzrA fusion to GFP were imaged by SIM every 5 s in the absence (EzrA control) or presence (EzrA + PC) of PC190723. Kymographs were obtained by extracting fluorescence intensity values along the red line indicated in cells in the left panels. Similar to the observations for $FtsZ^{55–56}$sGFP, the addition of PC190723 abolished EzrA movement (vertical lines in the kymographs). **b**, ColpSGEzrA-GFP cells were imaged by SIM every 5 min in the absence (left) or presence (right) of PC190723, and kymographs that show the constriction of EzrA–GFP rings were plotted. Under control conditions, the larger EzrA rings showed biphasic constriction behaviour whereas in the presence of PC190723 only rings in the second stage of cytokinesis were able to constrict. Data in **a** and **b** are representative of two biological replicates. Scale bars, 1 μm. **c**, To test the functionality of the EzrA–GFP construct, the strains COL, ColpSGEzrA-GFP and COLΔEzrA (which lacks *ezrA*) were imaged by phase contrast and cell areas were measured. The lack of EzrA in COLΔEzrA ($n = 959$) resulted in cell enlargement, whereas the size distribution of ColpSGEzrA-GFP ($n = 957$) cells mimicked that of parental strain COL ($n = 851$), which indicates that the EzrA fluorescent fusion is functional.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



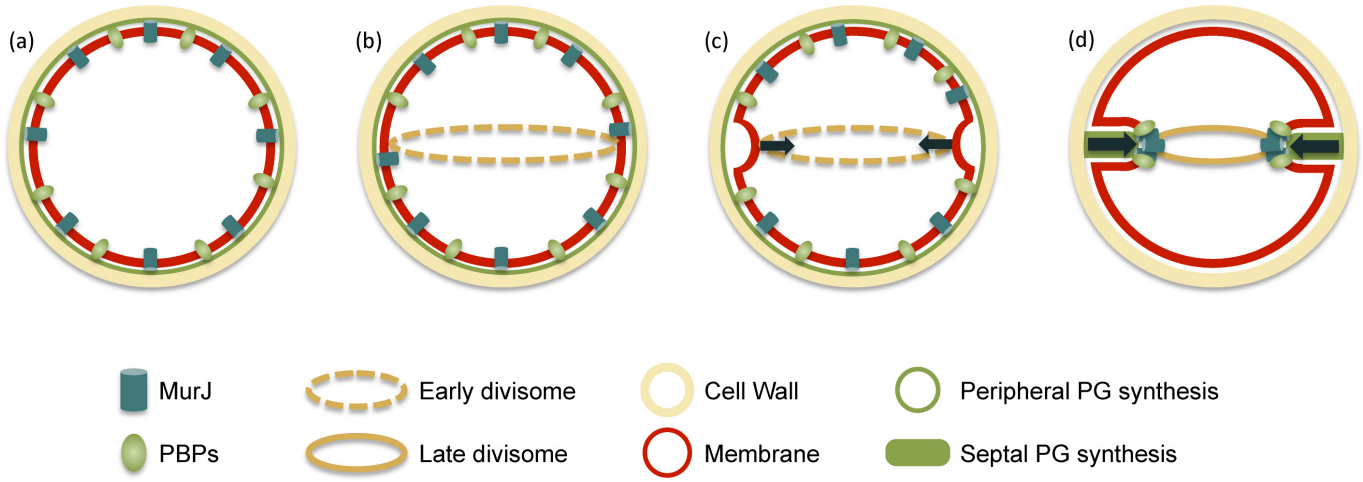
Extended Data Figure 8 | Kymographs that show the constriction of FtsZ$^{55–56}$sGFP, FtsW–sGFP and MurJ–sGFP rings during cell division. The strains ColFtsZ$^{55–56}$sGFP, ColFtsW-sGFP and ColMurJ-sGFP were imaged every 10 min in the absence (control) or presence (+PC) of PC190723 for a total of 60 min. MurJ–sGFP control kymographs were performed on cells in which the MurJ–sGFP signal appeared on the second frame to ensure that the entire constriction process was observed (that is, to confirm that the absence of a biphasic behaviour was not the result of only imaging cells in the later stages of cell division). FtsZ$^{55–56}$sGFP and FtsW–sGFP rings showed biphasic constriction behaviour (no or slow constriction, followed by fast constriction). The addition of PC190723 inhibited the constriction of larger FtsZ$^{55–56}$sGFP and FtsW–sGFP rings (top kymographs) but not of smaller rings that were undergoing fast constriction. MurJ–sGFP rings displayed only fast constriction, and therefore were always able to constrict in the presence of PC190723. Data are representative of three biological replicates. Scale bars, 0.5 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 9 | Graphs of FtsZ, MurJ and FtsW ring diameter during constriction. a, Graphs of ring diameters of FtsZ–mCherry (red) and MurJ–sGFP (blue) in strain ColJgZm. SIM images were taken every 10 min and measurements of ring diameter of FtsZ–mCherry and MurJ–sGFP were performed in cells in which a MurJ–sGFP signal first appeared after 20 min. **b**, Graphs of ring diameters of FtsZ–mCherry (red) and FtsW–sGFP (blue) in the strain ColWgZm. SIM images were taken every 10 min and measurements of ring diameter of FtsZ–mCherry and FtsW–sGFP were performed in cells in which FtsZ–mCherry ring diameter was constant for at least the first 20 min.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | Model for cytokinesis in *S. aureus*. **a**, Cells in phase 1 of the cell cycle (before septum synthesis is initiated) synthesize peptidoglycan at the cell periphery. **b**, In preparation for division, FtsZ and early components of the divisome assemble at the midcell. **c**, At this stage, cells initiate the first, slow step of cytokinesis (which is dependent on FtsZ treadmilling) and FtsZ may act as the driving force for the initial invagination of the membrane. **d**, MurJ then arrives at the divisome, in a process dependent on the presence of the sub-complex DivIB–DivIC–FtsL, bringing the PG precursor (lipid II) flippase activity to the midcell. The major *S. aureus* PG synthase (PBP2) is recruited to the midcell through substrate (translocated lipid II) recognition, and massive PG synthesis is initiated to synthesize the septum. From this point on, cytokinesis no longer depends on FtsZ treadmilling and is most probably driven by PG synthesis.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature25751

Transcriptional regulation by NR5A2 links differentiation and inflammation in the pancreas

Isidoro Cobo[1], Paola Martinelli[1]†*, Marta Flández[1]†*, Latifa Bakiri[2], Mingfeng Zhang[3], Enrique Carrillo-de-Santa-Pau[1], Jinping Jia[3], Víctor J. Sánchez-Arévalo Lobo[1]†, Diego Megías[4], Irene Felipe[1], Natalia del Pozo[1,5], Irene Millán[1,5], Liv Thommesen[6], Torunn Bruland[7,8], Sara H. Olson[9], Jill Smith[10], Kristina Schoonjans[11], William R. Bamlet[12], Gloria M. Petersen[12], Núria Malats[13,14], Laufey T. Amundadottir[3], Erwin F. Wagner[2] & Francisco X. Real[1,5,15]

Chronic inflammation increases the risk of developing one of several types of cancer. Inflammatory responses are currently thought to be controlled by mechanisms that rely on transcriptional networks that are distinct from those involved in cell differentiation[1–3]. The orphan nuclear receptor NR5A2 participates in a wide variety of processes, including cholesterol and glucose metabolism in the liver, resolution of endoplasmic reticulum stress, intestinal glucocorticoid production, pancreatic development and acinar differentiation[4–8]. In genome-wide association studies[9,10], single nucleotide polymorphisms in the vicinity of *NR5A2* have previously been associated with the risk of pancreatic adenocarcinoma. In mice, *Nr5a2* heterozygosity sensitizes the pancreas to damage, impairs regeneration and cooperates with mutant *Kras* in tumour progression[11]. Here, using a global transcriptomic analysis, we describe an epithelial-cell-autonomous basal pre-inflammatory state in the pancreas of *Nr5a2*$^{+/-}$ mice that is reminiscent of the early stages of pancreatitis-induced inflammation and is conserved in histologically normal human pancreases with reduced expression of *NR5A2* mRNA. In *Nr5a2*$^{+/-}$ mice, NR5A2 undergoes a marked transcriptional switch, relocating from differentiation-specific to inflammatory genes and thereby promoting gene transcription that is dependent on the AP-1 transcription factor. Pancreatic deletion of *Jun* rescues the pre-inflammatory phenotype, as well as binding of NR5A2 to inflammatory gene promoters and the defective regenerative response to damage. These findings support the notion that, in the pancreas, the transcriptional networks involved in differentiation-specific functions also suppress inflammatory programmes. Under conditions of genetic or environmental constraint, these networks can be subverted to foster inflammation.

The relationship between germline *Nr5a2* haploinsufficiency, mutant *Kras* and inflammation in promoting pancreatic adenocarcinoma (PDAC) suggests functional interactions that may be relevant to human disease[11]. In two independent groups of patients with PDAC, tumours with lower levels of NR5A2 expression were significantly enriched in patients with a history of chronic pancreatitis ($P = 0.001$) (Supplementary Table 1). The rs3790844 risk-increasing allele (T), which is associated with increased susceptibility to pancreatic cancer[9], is also associated with reduced NR5A2 protein levels ($P = 0.028$, $\beta = -0.57$) (Extended Data Fig. 1). This suggests that NR5A2 expression may underlie susceptibility to PDAC.

We used *Nr5a2*$^{+/-}$ mice to identify the mechanisms that favour impaired regeneration in the pancreas and PDAC. The pancreas of an adult *Nr5a2*$^{+/-}$ mouse is histologically normal, as are its expression of pancreatic transcription factors and digestive enzymes (Extended Data Fig. 2a, b). RNA sequencing (RNA-seq) showed that 926 and 100 genes were significantly upregulated or downregulated, respectively, in pancreases from 8–10-week-old *Nr5a2*$^{+/-}$ mice, when compared to wild-type mice. Gene set enrichment analysis (GSEA) revealed that 21 out of 23 gene sets that were over-represented in *Nr5a2*$^{+/-}$ mice, and 68% of upregulated genes, belong to inflammatory pathways (Fig. 1a, b and Supplementary Tables 2, 3). Among these genes are those coding for chemokines (*Ccl5* and *Ccl7*), complement components (*C1qb* and *C3*) and metalloproteases. Results were confirmed by quantitative PCR (qPCR) with reverse transcription (qRT–PCR) (Fig. 1c). Corresponding changes in H3K27ac or H3K27me3 at the promoters of differentially regulated genes were also observed (data not shown). mRNA and protein analyses using fresh tissue, isolated acini and immunohistochemistry confirmed upregulation of inflammatory mediators in acinar cells (Fig. 1c–e and Extended Data Figs 2c, 3a). Quantification of Cd45^{+} cells and immune-cell populations in wild-type and *Nr5a2*$^{+/-}$ pancreases did not reveal significant differences between the two, and a role for a ductal cell contribution to inflammatory gene expression was excluded (Fig. 1f and Extended Data Figs 2d–f, 3). Pancreas-selective inactivation of one *Nr5a2* allele, and biallelic *Nr5a2* inactivation in myeloid cells, showed that the defect is specific to pancreatic epithelial cells (Extended Data Fig. 4), which indicates that *Nr5a2* haploinsufficiency generates an epithelial-cell-autonomous basal pre-inflammatory state.

To assess whether similar transcriptomic changes occur in humans, we compared histologically normal pancreases ($n = 95$)[12] with high versus low *NR5A2* expression (top versus bottom quartile) ($n = 24$ per group). Seventy-eight per cent of genes for which expression was upregulated in *Nr5a2*$^{+/-}$ pancreases ($n = 718$) were also differentially expressed in human samples ($P < 0.05$). Of these differentially upregulated genes, 92% were significantly upregulated. By contrast, 44% of downregulated genes ($n = 64$) were differentially represented ($P < 0.05$) and only 50% of them were downregulated (Fig. 1g). Overall, genes that were upregulated in *Nr5a2*$^{+/-}$ mice were also upregulated

[1]Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre-CNIO, Madrid, Spain. [2]Genes, Development, and Disease Group, Spanish National Cancer Research Centre-CNIO, Madrid, Spain. [3]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [4]Confocal Microscopy Unit, Spanish National Cancer Research Centre-CNIO, Madrid, Spain. [5]Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre-CNIO, CIBERONC, Madrid, Spain. [6]Department of Biomedical Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [7]Clinic of Medicine, St. Olav's University Hospital, Trondheim, Norway. [8]Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [9]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [10]Departments of Gastroenterology and Hepatology, Georgetown University, Washington, DC 20007, USA. [11]Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [12]Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota 55905, USA. [13]Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre-CNIO, Madrid, Spain. [14]Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre-CNIO, CIBERONC, Madrid, Spain.[15]Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain. †Present addresses: Medical University Vienna, Department of Medicine I, Institute for Cancer Research, Vienna, Austria (P.M.); Tissue Regeneration Laboratory, Spanish National Cardiovascular Research Centre-CNIC, Madrid, Spain (M.F.); Department of Biotechnology, Facultad de Ciencias Biosanitarias, Universidad Francisco de Vitoria, Pozuelo de Alarcón, Spain (V.J.S.-A.L.).
*These authors contributed equally to this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

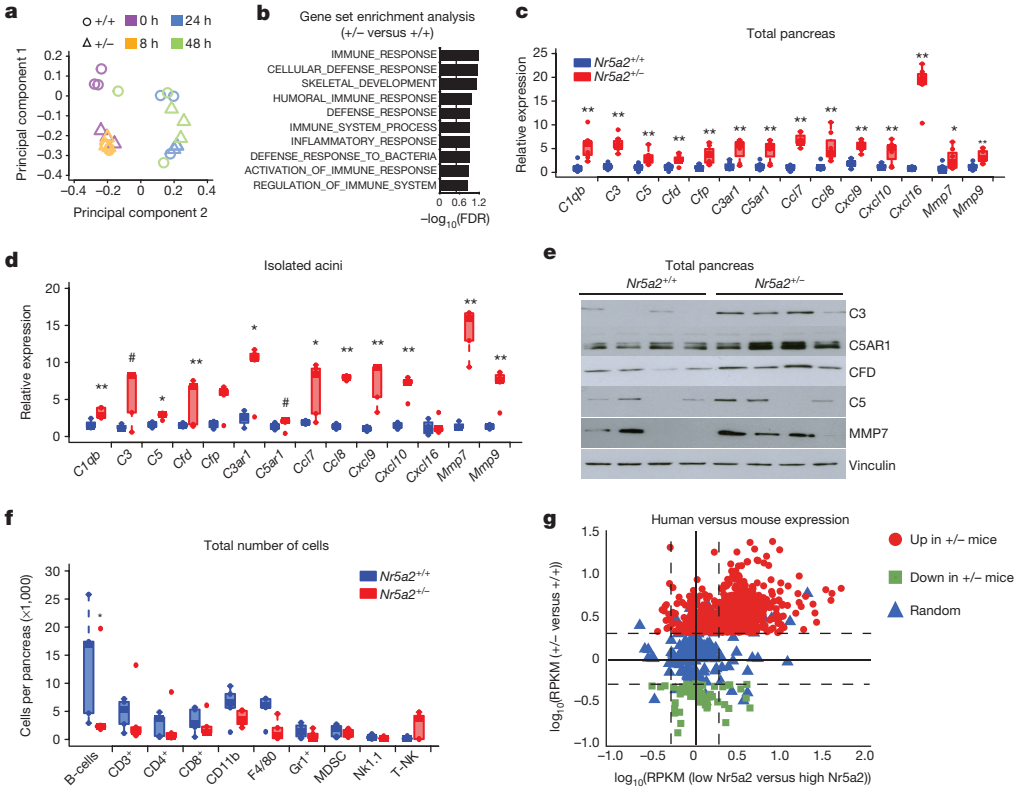


Figure 1 | Reduced NR5A2 expression is associated with pre-inflammation in normal pancreas of mice and humans. a, Principal component analysis of the transcriptome of wild-type (+/+) and *Nr5a2*$^{+/-}$ (+/−) mice in basal conditions and during caerulein-induced pancreatitis (seven injections, given once an hour) ($n = 3$ per group). **b**, GSEA of genes that are differentially expressed in *Nr5a2*$^{+/-}$ ($n = 3$) and wild-type mice ($n = 3$). **c**, qRT–PCR showing upregulation of inflammatory genes in *Nr5a2*$^{+/-}$ pancreases relative to wild type ($n = 7$ per group). **d**, qRT–PCR showing upregulation of inflammatory genes in primary *Nr5a2*$^{+/-}$ acinar cells relative to wild type ($n = 4$ per group). **e**, Western blots showing upregulation of inflammatory proteins in *Nr5a2*$^{+/-}$ pancreases ($n \geq 4$ per group). **f**, Fluorescence-activated cell-sorting analysis of inflammatory cells in wild-type and *Nr5a2*$^{+/-}$ pancreases ($n = 4$ per group). **g**, Scatter plot showing the relationship between expression of upregulated, downregulated or a random set of genes in *Nr5a2*$^{+/-}$ versus control *Nr5a2*$^{+/+}$ mice (*y* axis) and in histologically normal human pancreatic samples (*x* axis, low versus top high quartiles of *NR5A2* mRNA expression). In **c**, **d** and **f**, one-sided Mann–Whitney *U* test; $*P < 0.05$, $**P < 0.01$.

in NR5A2^{low} versus NR5A2^{high} human pancreases, when compared with a random list of 173 genes ($P = 1 \times 10^{-31}$). By contrast, downregulated genes did not follow a concordant pattern in mouse and human pancreases ($P = 0.58$). The conservation of findings across species provides a framework for understanding the relationship between the *NR5A2* genotype, *NR5A2* mRNA expression and PDAC in mice.

Analysis of public ChIP–seq (chromatin immunoprecipitation followed by sequencing) datasets[6] pertaining to pancreatic NR5A2 showed that the promoters of 89% of the genes that had upregulated expression in *Nr5a2*$^{+/-}$ mice contain putative NR5A2 binding sites, but only 7% actually exhibit NR5A2 binding within 2.5 Kb of the transcriptional start site. This suggests the participation of an indirect mechanism (Fig. 2a). Promoter scanning analysis of upregulated genes showed significant enrichment of AP-1 and NF-κB binding motifs, among others. A similar enrichment was found when computing the list of upregulated genes with GSEA using the MSigDB C3 transcription factor target gene-set collection (http://software.broadinstitute.org/gsea/msigdb/collections.jsp). *Jun* (also known as *c-Jun*), *Junb*, *Jund* and *Fos* (also known as *c-Fos*) mRNA and their corresponding proteins were significantly upregulated in *Nr5a2*$^{+/-}$ pancreases (Fig. 2b, c). We confirmed that these changes occurred in epithelial cells using freshly isolated acini and immunohistochemistry (Extended Data Fig. 3g). Phospho-JUN (p-JUN) and phospho-JNK (p-JNK) were also upregulated in *Nr5a2*$^{+/-}$ pancreases and isolated acini (Fig. 2d and Extended Data Fig. 3h). ChIP followed by qPCR (ChIP–qPCR) showed significantly increased binding of NR5A2 to the promoters of *Fos* and *Fosl1* (also known as *Fra1*)—but not to those of *Jun*, *Junb* and *Jund*—in *Nr5a2*$^{+/-}$ pancreases compared to wild-type mouse pancreases (Fig. 2e). NR5A2 expression in human embryonic kidney (HEK293) cells led to a dose-dependent decrease of *JUN* mRNA (Fig. 2f), indicating that changes in NR5A2 levels can modulate JUN expression. ChIP–qPCR showed increased binding of JUN, JUNB, JUND and FOS to the promoter of upregulated inflammatory genes in *Nr5a2*$^{+/-}$ pancreases (Fig. 2g).

The basal pre-inflammatory transcriptome of *Nr5a2*$^{+/-}$ mice suggests a subclinical pancreatitis-like state. We compared the transcriptomes of wild-type and *Nr5a2*$^{+/-}$ pancreases 8, 24 and 48 h after induction of mild acute pancreatitis (with 7 doses of caerulein), using RNA-seq. Principal component analysis highlighted the divergence of the basal transcriptomes of wild-type and *Nr5a2*$^{+/-}$ pancreases; these differences were completely but transiently eroded 8 h after pancreatitis induction. The basal transcriptome of *Nr5a2*$^{+/-}$ pancreases was similar to that of wild-type mouse pancreases at 8 h (Fig. 1a). The differences between the set of genes upregulated in *Nr5a2*$^{+/-}$ pancreases under basal conditions and those upregulated in wild-type pancreases were notably reduced 8–24 h after pancreatitis induction; a similar pattern of reduced differences between *Nr5a2*$^{+/-}$ and wild-type pancreases was observed for the downregulated genes. The expression of a random set of genes was unaffected (Extended Data Fig. 5a). Consistently, although there was a strong upregulation of AP-1 at 8 h after induction of pancreatitis in mice of both genotypes, persistent AP-1 upregulation at 24–48 h was observed only in *Nr5a2*$^{+/-}$ mice (Extended Data Fig. 5b). These results indicate that *Nr5a2* haploinsufficiency mimics acute pancreatitis at the transcriptome level.

The repeated caerulein dosing required to induce pancreatitis hampers the interpretation of dynamic acute signalling and transcriptional responses. Therefore, we also analysed the effects of a single dose of caerulein that did not induce pancreatitis (Extended Data Fig. 6a). JUN,

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

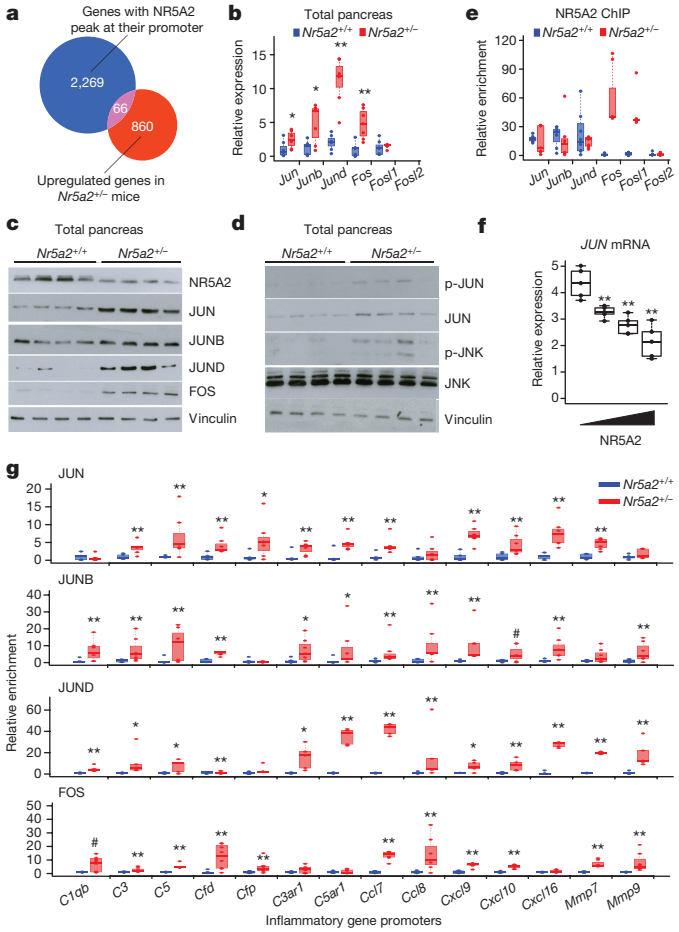


Figure 2 | AP-1 components are upregulated and bind to the promoter of inflammatory genes in *Nr5a2*$^{+/-}$ pancreases. a, A total of 926 genes are upregulated in *Nr5a2*$^{+/-}$ pancreases. In ChIP–seq experiments using normal pancreas cells, 2,269 genes have an NR5A2 peak at their promoter. A minor fraction of the promoters of genes upregulated in *Nr5a2*$^{+/-}$ pancreases (66 out of 926) also display NR5A2 peaks at their promoters in normal pancreas cells. **b**, qRT–PCR of AP-1 component expression in an independent group of wild-type and *Nr5a2*$^{+/-}$ mice ($n = 7$ per group). **c**, **d**, Western blots showing NR5A2, AP-1, p- JUN and p-JNK expression in wild-type and *Nr5a2*$^{+/-}$ pancreases ($n = 4$ per group). **e**, NR5A2 enrichment at AP-1 gene promoters revealed by ChIP–qPCR ($n = 7$ per group). **f**, qRT–PCR of JUN expression in HEK293 cells upon transfection of a plasmid coding for NR5A2, showing reduced *Jun* mRNA (mean of five independent experiments). **g**, ChIP–qPCR of the occupancy of inflammatory gene promoters by AP-1 (JUN, JUNB, JUND and FOS) in wild-type and *Nr5a2*$^{+/-}$ pancreases ($n \geq 5$ per group). At least two independent experiments were performed in all cases. In **b**, **e**, **f** and **g**, one-sided Mann–Whitney *U* test; $*P < 0.05$, $**P < 0.01$.

FOS, JUND, FOSL1 and FOSL2 (also known as FRA2) were upregulated to a similar extent in both wild-type and *Nr5a2*$^{+/-}$ pancreases 30–60 min after caerulein administration, preceding STAT3 phosphorylation (Extended Data Fig. 6b). We then analysed the expression of a subset of the genes differentially overexpressed in *Nr5a2*$^{+/-}$ pancreases under basal conditions. By 30–60 min, one dose of caerulein was sufficient to promote an inflammatory profile in wild-type mice; this profile closely resembled that of *Nr5a2*$^{+/-}$ pancreases. A complete resolution of this expression profile occurred in wild-type mice by 12 h (Fig. 3a, b). To determine whether these are epithelial-cell-autonomous effects, primary acini from wild-type and *Nr5a2*$^{+/-}$ mice were treated with vehicle or caerulein for 24 h. qRT–PCR analysis confirmed that under control conditions *Nr5a2*$^{+/-}$ acini express higher levels of inflammatory genes than wild-type acini, and that these differences were abolished upon caerulein treatment (Extended Data Fig. 6c). These findings further support the notion that *Nr5a2*$^{+/-}$ acini display a transcriptional pre-inflammatory phenotype similar to that of caerulein-treated wild-type acini.

A large fraction of inflammatory gene promoters contain putative NR5A2 binding sites that are not occupied in normal pancreas (Fig. 2a, data not shown). Accordingly, we performed ChIP–qPCR on the promoter of bona fide NR5A2 target differentiation-related genes (that is, those encoding digestive enzymes) and of inflammatory genes upregulated in *Nr5a2*$^{+/-}$ mice. In untreated wild-type mice NR5A2 was observed to bind at the promoter of differentiation genes, but not at the promoter of inflammatory genes (Fig. 3c). However, in untreated *Nr5a2*$^{+/-}$ mice, the binding of NR5A2 to differentiation gene promoters was reduced and binding to inflammatory gene promoters was increased. We designate this relocation of NR5A2 between gene sets as the 'NR5A2 transcriptional switch'. A similar transcriptional switch occurred in wild-type mice 30 min after 1 dose of caerulein and the genomic distribution of NR5A2 was restored 12 h later (Fig. 3d), supporting its physiological relevance. NR5A2 can interact with JUN *in vitro*[13], and we found that both proteins co-immunoprecipitated in the pancreases of untreated *Nr5a2*$^{+/-}$ mice and in wild-type mice 1 h after caerulein administration, but not in untreated wild-type mice (Fig. 3e).Using sequential ChIP–qPCR, NR5A2 and JUN were detected on the same immunoprecipitated chromatin regions at the promoters of *C1qb*, *Ccl7* and *Ccl8* in untreated *Nr5a2*$^{+/-}$ mice and in wild-type mice 1 h after caerulein administration, supporting the proposition that JUN and AP-1 cooperate in producing the NR5A2 transcriptional switch (Fig. 3f and Extended Data Fig. 7).

NR0B2 is an NR5A2 co-repressor and target gene[14] that is highly expressed in acinar cells, shows reduced expression in *Nr5a2*$^{+/-}$ pancreases and is dynamically regulated on caerulein administration (Extended Data Fig. 8a–e). In humans, *NR0B2* mRNA expression was significantly higher in NR5A2^{high} pancreases ($P = 1.4 \times 10^{-7}$) and NR5A2 and NR0B2 levels were positively correlated with one another ($R^2 = 0.312$) (Extended Data Fig. 8f), which suggests that NR0B2 contributes to the haploinsufficient phenotype. Under basal conditions in wild-type mice, NR0B2 was absent from the promoter of differentiation and inflammatory genes but it was enriched in the promoter of *Jun*, *Junb* and *Jund* (Extended Data Fig. 8g). *Nr5a2*$^{+/-}$ pancreases displayed reduced NR0B2 levels, lower occupancy of AP-1 promoters by NR0B2, reduced binding of NR5A2 at the *Nr0b2* promoter and reduced complexing of NR5A2 with NR0B2 (Extended Data Fig. 8c, g–i), which suggests that NR0B2 ordinarily represses AP-1 expression. We validated the 266-6 mouse acinar cell line as a model to assess the effects of caerulein on AP-1 and inflammatory genes (Extended Data Fig. 9a–c). On expression of NR0B2, *Jun* mRNA—but no inflammatory transcript—was reduced and NR0B2 knockdown modestly increased NR5A2 occupancy of inflammatory gene promoters (Extended Data Figs 9d, e). The effects of NR5A2 knockdown on JUN and inflammatory gene expression were abrogated by NR0B2 overexpression (Extended Data Fig. 9f), supporting the notion that NR5A2 modulates AP-1 expression in part through NR0B2 downregulation.

AP-1 upregulation, increased binding to inflammatory gene promoters in *Nr5a2*$^{+/-}$ mice and the interaction of JUN and NR5A2 all suggest a crucial involvement of JUN in the basal pre-inflammatory state. *Jun* deletion in *Nr5a2*$^{+/-}$ pancreases rescued the upregulation of AP-1 and inflammatory transcripts, as well as NR5A2 binding to AP-1 and inflammatory gene promoters, without affecting NR5A2 expression (Fig. 4a–d). On induction of acute pancreatitis, *Nr5a2*$^{+/-}$ mice in which the *Jun* gene has been deleted in the pancreas (*Nr5a2*$^{+/-}$;*Jun*ΔP) showed reduced damage when compared with control *Nr5a2*$^{+/-}$ mice (Fig. 4e, f). The normalization of the dynamic expression of AP-1 components during pancreatitis in *Nr5a2*$^{+/-}$;*Jun*ΔP pancreases suggests a critical role for JUN in the phenotype of *Nr5a2*$^{+/-}$ mice. Immunohistochemical analysis showed that the downregulation of JUN occurred selectively in acinar cells (Fig. 4g and Extended Data Fig. 10). Taken together, these results indicate that *Jun* is required to generate the pre-inflammatory state observed in *Nr5a2*$^{+/-}$ pancreases.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

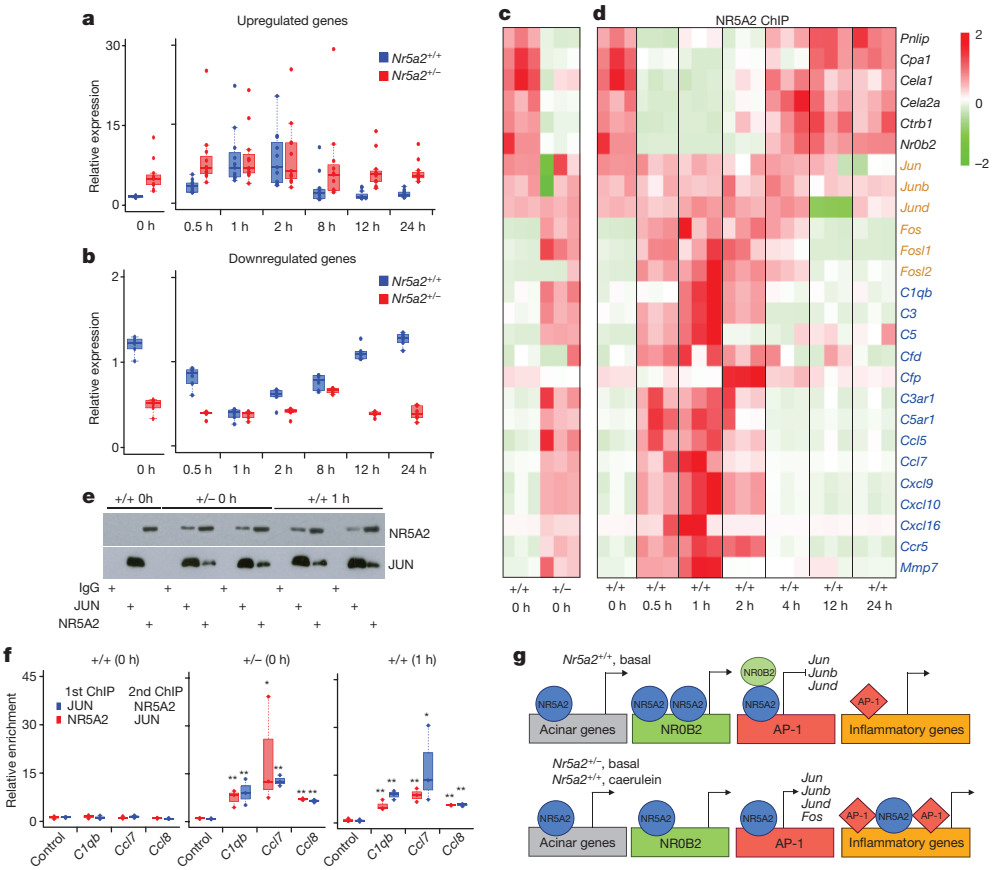


Figure 3 | The NR5A2 transcriptional switch. a, **b**, Dynamic gene expression ($Nr5a2^{+/-}$ versus $Nr5a2^{+/+}$) after one dose of caerulein, shown using qRT–PCR. Data refer to basal values in wild-type pancreases ($n = 4$ per group; two independent experiments). **c**, ChIP–qPCR showing differential promoter occupancy by NR5A2 under basal conditions ($n = 3$ per group). **d**, ChIP–qPCR showing NR5A2 switching from the promoter of pancreatic genes to the promoter of inflammatory genes after one caerulein dose in wild-type mice (black, pancreatic genes; orange, AP-1 genes; blue, inflammatory genes) ($n = 3$ per group; two independent experiments). **e**, Co-immunoprecipitation and western blot showing that NR5A2 and JUN are part of the same complex in $Nr5a2^{+/-}$ pancreases under basal conditions and in wild-type mouse pancreases after one dose of caerulein, but not in wild-type mouse pancreases under basal conditions (at least two independent experiments). **f**, Sequential ChIP–qPCR of JUN–NR5A2 and NR5A2–JUN to assess promoter binding in wild-type pancreases under basal conditions, and 1 h after caerulein dosing (see Extended Data Fig. 7) (n = 3 pools of 3 pancreases). One-sided Mann–Whitney *U* test; $^*P < 0.05$, $^{**}P < 0.01$. **g**, Proposed model of the NR5A2 transcriptional switch. Under basal conditions in wild-type mice, NR5A2 is bound to acinar gene promoters as well as to the promoter of *Nr0b2*. In this condition, AP-1 proteins show weak binding to inflammatory gene promoters. By contrast, in untreated $Nr5a2^{+/-}$ mice, and in wild-type mice that have received caerulein, NR5A2 is weakly bound to the promoter of acinar genes but shows increased binding to the promoters of *Fos*, *Fosl1* and inflammatory genes; AP-1 proteins also show increased binding to these promoters. In wild-type mice dosed with caerulein, there is a transient redistribution of NR5A2 to AP-1 genes, and of AP-1 proteins to inflammatory genes. NR0B2 levels are reduced in untreated $Nr5a2^{+/-}$ mice and in wild-type mice that have received caerulein. This is accompanied by reduced binding to AP-1 promoters, which suggests that NR0B2 participates in the regulation of AP-1 gene expression.

In homeostatic conditions, tissue inflammation is suppressed. However, a direct link between differentiation and inflammatory programmes has not yet been proposed. Here, we demonstrate that NR5A2 is critical in restraining inflammation in normal mouse pancreas cells. Constitutive loss of one *Nr5a2* allele produces a pre-inflammatory state that explains how haploinsufficiency primes tissue for damage and delays recovery, which may contribute to accelerated tumorigenesis[11]. This pre-inflammatory state is mediated by JUN phosphorylation and stabilization, which are hallmarks of JUN activation[15]. In $Nr5a2^{+/-}$ mice, low NR5A2 and NR0B2 expression probably determine AP-1 dysregulation, which contributes to inflammatory gene expression, as occurs on induction of pancreatitis, (Fig. 3g). In the pancreas, AP-1 is pro-inflammatory; this is unlike the situation found in the skin, in which *Jun* and *Junb* deletion in keratinocytes induces inflammatory phenotypes[16]. Our results, therefore, support the notion that the effects of AP-1 are context-dependent and tissue-specific[17,18]. The output of the interaction of NR5A2, NR0B2 and AP-1 with gene regulatory elements is likely to be modulated by post-transcriptional modifications and chromatin accessibility: in keratinocytes, the coordinated regulation of polycomb-group and AP-1 proteins ensures canonical expression of epidermal genes during differentiation[19].

We find chemokines and complement components among the inflammatory genes most prominently upregulated in $Nr5a2^{+/-}$ pancreases. The control of inflammation from epithelial cells at the level of chemotactic stimuli suggests that leukocyte migration to tissues generates the local upregulation of cytokines, such as IL1, IL6 and TNF, which then contribute to amplifying and prolonging the inflammatory response[20]. The findings in $Nr5a2^{+/-}$ pancreases indicate that inflammation is actively repressed in normal epithelial cells, and that the genetic pre-inflammatory phenotype mimics the molecular events associated with response to pharmacological (such as caerulein) and environmental stimuli. Indeed, both conditions are associated with a switch in the chromatin distribution of NR5A2, which shifts from the promoters of pancreatic differentiation genes to those of inflammatory genes. Our findings in mice are relevant to human pancreatitis and PDAC, as shown by shared transcriptomic changes that indicate a pre-inflammatory state in the pancreas of human subjects with low *NR5A2* mRNA levels and by the association of this pre-inflammatory state with low levels of NR5A2 and NR0B2. Although additional work is needed to establish the functional consequences of carrying PDAC-risk alleles at chr1q32.1, our results suggest that the underlying biology at this locus may involve negative regulation of *NR5A2* expression.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

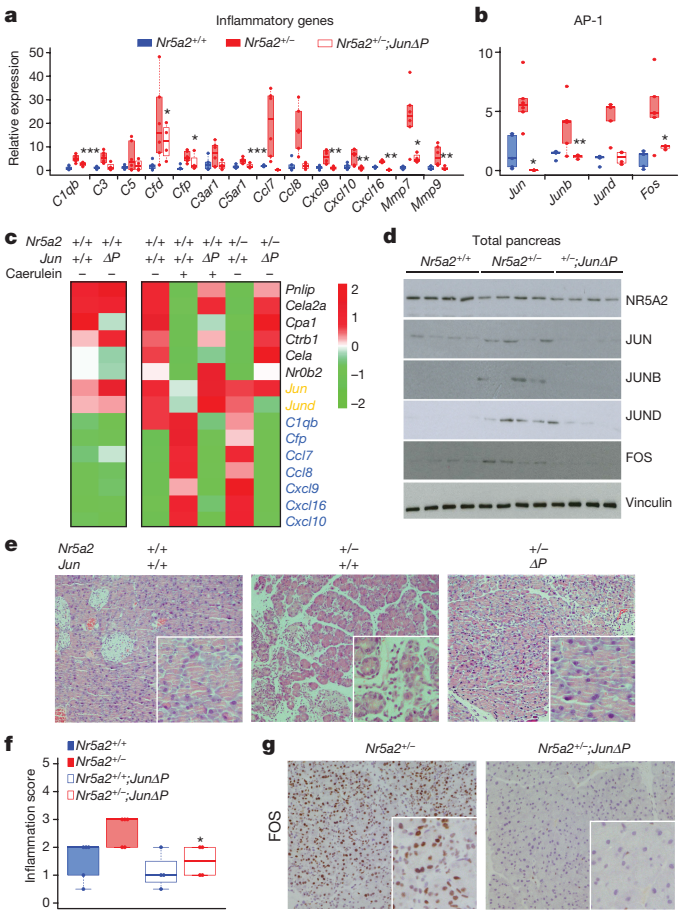


Figure 4 | Pancreatic deletion of *Jun* rescues the pancreatic defect of *Nr5a2*$^{+/-}$ mice. **a**, **b**, qRT–PCR analysis of inflammatory gene (**a**) and AP-1 (**b**) transcripts in the pancreas of wild-type, *Nr5a2*$^{+/-}$ and *Nr5a2*$^{+/-}$;*JunΔP* mice; **a** and **b** are shown relative to basal values in wild-type mice ($n \geq 4$). **c**, ChIP–qPCR shows that the NR5A2 transcriptional switch is reversed on the deletion of *Jun* in the pancreas ($n =$ pool of 3 pancreases). This experiment was performed once. **d**, Western blots showing the expression of NR5A2 and AP-1 in *Nr5a2*$^{+/+}$, *Nr5a2*$^{+/-}$ and *Nr5a2*$^{+/-}$;*JunΔP* pancreases ($n = 4$ per group). **e**, Histology of *Nr5a2*$^{+/+}$, *Nr5a2*$^{+/-}$ and *Nr5a2*$^{+/-}$;*JunΔP* pancreases 48 h after induction of acute pancreatitis shows that *Jun* deletion rescues the excessive damage caused by *Nr5a2* haploinsufficiency ($n \geq 4$ per group). All experimental data were obtained in a single experiment. **f**, Inflammation scores corresponding to the experiment in **e** ($n \geq 4$ per group), plus scores from *Nr5a2*$^{+/+}$ mice with a *Jun* deletion. **g**, Immunohistochemistry of FOS in *Nr5a2*$^{+/-}$ and *Nr5a2*$^{+/-}$;*JunΔP* pancreases 48 h after induction of acute pancreatitis ($n = 4$ per group; representative results of one pancreas). In **a**, **b** and **f**, one-sided Mann–Whitney *U* test; $^{*}P < 0.05$, $^{**}P < 0.01$.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 July 2016; accepted 15 January 2018.
Published online 14 February 2018.

1. Karin, M. & Clevers, H. Reparative inflammation takes charge of tissue regeneration. *Nature* **529,** 307–315 (2016).
2. Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140,** 883–899 (2010).
3. Crusz, S. M. & Balkwill, F. R. Inflammation and cancer: advances and new agents. *Nat. Rev. Clin. Oncol.* **12,** 584–596 (2015).
4. Stein, S. & Schoonjans, K. Molecular basis for the regulation of the nuclear receptor LRH-1. *Curr. Opin. Cell Biol.* **33,** 26–34 (2015).
5. Mamrosh, J. L. *et al.* Nuclear receptor LRH-1/NR5A2 is required and targetable for liver endoplasmic reticulum stress resolution. *eLife* **3,** e01694 (2014).
6. Holmstrom, S. R. *et al.* LRH-1 and PTF1-L coregulate an exocrine pancreas-specific transcriptional network for digestive function. *Genes Dev.* **25,** 1674–1679 (2011).
7. Molero, X. *et al.* Gene expression dynamics after murine pancreatitis unveils novel roles for Hnf1α in acinar cell homeostasis. *Gut* **61,** 1187–1196 (2012).
8. Hale, M. A. *et al.* The nuclear hormone receptor family member NR5A2 controls aspects of multipotent progenitor cell formation and acinar differentiation during pancreatic organogenesis. *Development* **141,** 3123–3133 (2014).
9. Petersen, G. M. *et al.* A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat. Genet.* **42,** 224–228 (2010).
10. Amundadottir, L. T. Pancreatic cancer genetics. *Int. J. Biol. Sci.* **12,** 314–325 (2016).
11. Flández, M. *et al. Nr5a2* heterozygosity sensitises to, and cooperates with, inflammation in *KRas*^{G12V}-driven pancreatic tumourigenesis. *Gut* **63,** 647–655 (2014).
12. Zhang, M. *et al.* Characterizing *cis*-regulatory variation in the transcriptome of histologically normal and tumour-derived pancreatic tissues. *Gut* http://dx.doi.org/10.1136/gutjnl-2016-313146 (2017).
13. Huang, S. C., Lee, C. T. & Chung, B. C. Tumor necrosis factor suppresses NR5A2 activity and intestinal glucocorticoid synthesis to sustain chronic colitis. *Sci. Signal.* **7,** ra20 (2014).
14. Oiwa, A. *et al.* Synergistic regulation of the mouse orphan nuclear receptor SHP gene promoter by CLOCK-BMAL1 and LRH-1. *Biochem. Biophys. Res. Commun.* **353,** 895–901 (2007).
15. Papavassiliou, A. G., Chavrier, C. & Bohmann, D. Phosphorylation state and DNA-binding activity of c-Jun depend on the intracellular concentration of binding sites. *Proc. Natl Acad. Sci. USA* **89,** 11562–11565 (1992).
16. Schönthaler, H.B., Guinea-Viniegra, J. & Wagner, E. F. Targeting inflammation by modulating the Jun/AP-1 pathway. *Ann. Rheum. Dis.* **70,** i109–i112 (2011).
17. Shaulian, E. & Karin, M. AP-1 as a regulator of cell life and death. *Nat. Cell Biol.* **4,** E131–E136 (2002).
18. Eferl, R. & Wagner, E. F. AP-1: a double-edged sword in tumorigenesis. *Nat. Rev. Cancer* **3,** 859–868 (2003).
19. Ezhkova, E. *et al.* Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. *Cell* **136,** 1122–1135 (2009).
20. Headland, S. E. & Norling, L. V. The resolution of inflammation: principles and challenges. *Semin. Immunol.* **27,** 149–160 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank O. Domínguez, J. Herranz, T. Lobato, L. Martínez, and Y. Cecilia, as well as members of the CNIO core facilities, Epithelial Carcinogenesis Group, and Genes, Development and Disease Group; L. Montuenga, C. Rodríguez-Ortigosa, B. Bréant and cited investigators for providing antibodies; and E. Batlle and P. Muñoz-Cánoves for critical comments. This study used the high-performance computational capabilities of the Biowulf Linux cluster (https://hpc.nih.gov/). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, US National Institutes of Health (NIH), nor does mention of trade names, commercial products or organizations imply endorsement by the US government. This work was supported in part by grants SAF2011-29530 and SAF2015-70553-R from the Ministerio de Economía y Competitividad (co-funded by the ERDF-EU), RTICC from the Instituto de Salud Carlos III (RD12/0036/0034, RD12/0036/0050) and grants 256974 and 289737 from the European Union Seventh Framework Program to F.X.R.; grants BFU 2012-40230 and SAF2015-70857 from the Ministerio de Economía y Competitividad (co-funded by the ERDF-EU) and Worldwide Cancer Research (13-0216) to E.F.W.; grants PI12/00815 and PI1501573 from the Fondo de Investigaciones Sanitarias, Instituto de Salud Carlos III, Spain and EUPancreas COST Action BM1204 to N.M.; grant P30CA008748 from the US NIH, National Cancer Institute to S.H.O.; the Intramural Research Program of the NIH, National Cancer Institute; and Mayo Clinic SPORE in Pancreatic Cancer funded by National Cancer Institute grant P50 CA102701. L.T. and T.B. were supported by the Department of Technology, Norwegian University of Science and Technology, the Central Norway Regional Health Authority and by the European Science Foundation. P.M. and I.C. are recipients of Juan de la Cierva and Beca de Formación del Personal Investigador, respectively, from Ministerio de Economía y Competitividad. I.F. is the recipient of a 'Juegaterapia-Amigos del CNIO' Postdoctoral Fellowship. F.X.R. acknowledges the support of Asociación Española Contra el Cáncer.

Author Contributions I.C., M.F. and F.X.R. conceived the study. I.C., M.F. and N.d.P. performed animal experiments. I.C., E.C.-d.-S.-P., M.Z. and J.J. conducted bioinformatics analyses. I.C., V.J.S.-A.L. and I.F. conducted *in vitro* studies using mouse cells. S.H.O., J.S., W.R.B., G.M.P. and N.M. provided samples and information on human subjects. W.R.B., G.M.P., N.M. and L.T.A. designed and performed clinical studies, obtained samples and performed human data analysis. I.M., D.M., L.T. and T.B. were involved in data analysis. L.B., K.S. and E.F.W. provided reagents. P.M., L.B., L.T.A. and E.F.W. had critical input into experimental design, data analysis and interpretation. I.C. and F.X.R. wrote the manuscript with contributions of P.M., L.B., L.T.A. and E.F.W. F.X.R. supervised the overall conduct of the study. All authors read and approved the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to F.X.R. (preal@cnio.es).

Reviewer Information *Nature* thanks F. Greten, R. MacDonald and G. Natoli for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

No statistical methods were used to pretermine sample size. The investigators were not blinded to allocation during experiments but were blinded to outcome assessment.

Mice and experimental manipulations. The following mouse strains were used: *Nr5a2*$^{+/-}$ (ref. 21), conditional floxed *Jun*[22], conditional floxed *Nr5a2*[23], *Lys*^{cre} (ref. 24), and *Ptf1a*$^{cre/+}$ knock-in mice (a gift from C. V. Wright)[25]. Pancreas-specific *Nr5a2*-heterozygous and *Jun*-deficient mice were generated by crossing *Nr5a2*$^{lox/+}$ and *Jun*$^{lox/lox}$ mice with *Ptf1a*$^{cre/+}$ mice, respectively. All crosses were maintained in a predominantly C57BL/6 background. Littermate control mice were used. All experiments were performed using 8–14-week-old mice. All animal procedures were approved by local and regional ethics committees (Institutional Animal Care and Use Committee and Ethics Committee for Research and Animal Welfare, Instituto de Salud Carlos III) and performed according to the European Union guidelines. After a mouse was killed by cervical dislocation, the pancreas was removed quickly and placed in buffered formalin for histological analysis or homogenized in denaturing buffer (4 M guanidine thiocyanate, 0.1 M Trizma HCl pH 7.5, 1% 2-mercaptoethanol) for RNA extraction. In addition, a small piece was snap-frozen for protein isolation.

Mild acute pancreatitis was induced by seven injections, given once per hour, of the CCK analogue caerulein (Bachem) at 50 μg kg^{-1}. In brief, animals were weighed before beginning the procedure and caerulein was administered intraperitoneally. Mice were killed by cervical dislocation 8, 24 and 48 h after the first injection. For the single-injection protocol, mice received a single dose of caerulein (50 μg kg^{-1}) administered intraperitoneally. Mice were killed by cervical dislocation at the indicated time points.

The number of mice used in each experiment is shown in the legend of each figure. For most experiments, at least five mice per group were used; if different numbers of mice were included in each group, the lowest number is provided. No specific randomization method was used.

Acinar cell isolation. Acinar cells were isolated by collagenase digestion and maintained at 37 °C in Dulbecco's modified Eagle medium containing 2% bovine serum albumin and 10 mM 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid as previously described[26]. Acini were treated either with caerulein (100 pM) or with saline for 24 h before RNA and protein isolation.

Immunofluorescence and immunohistochemical analyses. Immunofluorescence and immunohistochemical analyses were performed using 3-μm sections of formalin-fixed paraffin-embedded tissues, unless otherwise indicated. After deparaffinization and rehydration, antigen retrieval was performed by boiling in citrate buffer pH 6 for 10 min. For immunofluorescence, the sections were incubated for 45 min at room temperature with 3% BSA, 0.1% Triton X-100–PBS and then with the primary antibody overnight at 4 °C. For double or triple immunofluorescence, the corresponding antibodies were added simultaneously and incubated overnight at 4 °C. Sections were then washed with 0.1% Triton–PBS, incubated with the appropriate fluorochrome-conjugated secondary antibody, and nuclei were counter-stained with DAPI. After washing with PBS, sections were mounted with Prolong Gold Antifade Reagent (Life Technology).

For immunofluorescence quantification of Cd45^{+} cells, fresh-frozen tissues were used. Sections were fixed with 4% paraformaldehyde for 10 min at room temperature with gentle rotation, washed with PBS and incubated with 3% BSA and 0.1% Triton–PBS for 45 min at room temperature, followed by overnight incubation with primary antibodies recognizing PTF1a (gift from B. Bréant, INSERM) and Cd45 (NB110-93609, Novus Biologicals). After washing with PBS, sections were incubated with the appropriate fluorochrome-conjugated secondary antibodies, and nuclei were counter-stained with DAPI. After washing with PBS, sections were mounted as described. Images were analysed with Definiens Developer XD v.2.5 software (Definiens): a script was developed in-house for single-cell segmentation. Afterwards, cells were classified into subtypes by calculating the nuclear and cytoplasmic intensity of the different markers. Owing to the low quantities of marker-positive cells, a picture gallery of all single positives was generated and false-positive images were manually discarded by one of the co-authors (D.M.) for maximum accuracy; this process was performed in a blinded fashion. There were no differences between wild-type and *Nr5a2*$^{+/-}$ mice, and only the raw uncurated results are shown in Extended Data Fig. 2e.

For immunohistochemical analyses, after antigen retrieval endogenous peroxidase was inactivated with 3% H_2O_2–methanol for 30 min at room temperature. Sections were incubated with 2% BSA–PBS for 1 h at room temperature, and incubated with the primary antibody. After washing, the Envision secondary reagent (DAKO) was added for 40 min at room temperature and sections were washed three times with PBS. 3,30-Diaminobenzidine tetrahydrochloride (DAB) was used as a chromogen. Sections were lightly counterstained with haematoxylin, dehydrated and then mounted. A non-related IgG was used as a negative control.

To identify ductal cells, the *Dolichos biflorus* agglutinin (DBA)–streptavidin visualization system was used (Discovery DAB Map Kit, RUO, Ventana). Sections were incubated with biotinylated DBA (1:1000; Vector Laboratories B-1035), washed and incubated with streptavidin–peroxidase. Reactions were developed using DAB. Histological images were acquired with a Nikon TE2000 microscope and the percentage of cells expressing DBA was quantified.

To quantify Cd45^{+} leukocytes using immunohistochemistry, an automated immunostaining platform was used (Ventana Discovery XT, Roche). Antigen retrieval was first performed with boiling citrate buffer pH 6.0; endogenous peroxidase was blocked and slides were incubated with rat monoclonal anti-CD45 (5C16, 1:500; Novus Biologicals, NB110-93609). Slides were incubated with the corresponding secondary antibodies (biotinylated rabbit anti-rat Ig, Vector Labs) and visualization systems (OmniRabbit, Ventana, Roche) conjugated with horseradish peroxidase. Immunohistochemical reactions were developed using DAB and nuclei were counterstained with haematoxylin. Finally, the slides were dehydrated, cleared and mounted for microscopic evaluation. Whole digital slides were acquired with a scanner (Axio Scan Z1, Zeiss) and images captured with Zen Software (Zeiss). Image analysis and quantification were performed using with the AxioVision software package (Zeiss). After selecting regions of interest, areas were selected for quantification and exported as subset TIFF images. Images with artefactual staining or cutting artefacts were eliminated. Afterwards, the images from all slides were chosen for automatic quantification (AxioVision 4.6, Zeiss) using a script for each antibody. For Cd45 quantification, positivity was evaluated in one phase (phase 1, positive) and compared with total tissue area (phase 2). The output results were then exported as Excel files with scoring data for each TIFF file. Data obtained were then compiled and appropriately assessed.

Histological scoring of mouse pancreases. Inflammation-related histological parameters (oedema, inflammatory cell infiltration and acino-ductal metaplasia) were scored blindly according to the grade of severity (0–3).

Flow cytometry analysis of inflammatory cells. In brief, mouse pancreases were injected *in situ* with collagenase P (1 mg/mL) (Roche), transferred to cold collagenase and minced. After incubating at 37 °C for 20 min with mild shaking, collagenase was inactivated with 5% cold FBS in HBSS and the pancreas was disaggregated and filtered twice through a 70-μM strainer. After centrifugation for 2 min at 300*g* and resuspension (2% chelated FBS, 2 mM EDTA, supplemented with DNase in HBSS) (Ambion), live cells were counted, incubated with blocking buffer and subsequently with primary antibody for 1 h at room temperature. Cell suspensions were analysed through a FACS ARIA IIu sorter coupled to a LSR Fortessa analyser with 10 different markers (Fig. 1f and Extended Data Fig. 2d). Mice with low counts of total live cells were not considered in the analyses. For each condition, $\geq$ 5 mice were used.

RNA-seq and data processing in mouse cells. Total pancreatic RNA was isolated using guanidine thiocyanate, followed by acid phenol–chloroform extraction. RNA integrity numbers ranged from 6.6 to 9.2, when assayed by laboratory chip technology on an Agilent 2100 Bioanalyzer. PolyA^{+} RNA was extracted and randomly fragmented, converted to double stranded cDNA and processed through subsequent enzymatic treatments of end-repair, dA-tailing and ligation to adapters according to Illumina's 'TruSeq RNA Sample Preparation Guide' (Part # 15008136 Rev. A; for samples Big104, Big90, Big92, Big18, Big277, Big278, Big33, Big87, Big94, Big113, Big17, Big86) or Illumina's 'TruSeq RNA Sample Preparation v.2 Protocol' (Part # 15026494 Rev. C; for samples Big408, Big409, Big410, Big416, Big417, Big423, Big454, Big459, Big461, Big465, Big467 and Big470). The adaptor-ligated library was completed by limited-cycle PCR with Illumina PE primers (10 cycles, or 8 cycles for samples using the v.2 protocol). The resulting purified cDNA library was applied to an Illumina flow cell for cluster generation (TruSeq cluster generation kit v.5) and sequenced on the Genome Analyzer IIx with SBS TruSeq v.5 reagents, according to the manufacturer's protocols. Three pancreases were analysed for each genotype (wild type or *Nr5a2*$^{+/-}$) and condition (baseline, 8 h, 24 h and 48 h after pancreatitis induction).

Image analysis and per-cycle base-calling were performed with Illumina real time analysis software (RTA1.13). Conversion to FASTQ read format with the ELAND algorithm (v.2e) was performed with CASAVA-1.8 (Illumina). These files contain only reads that passed 'chastity' filtering (tagged with an 'N', for *NOT filtered* in the sequence identifier line). Quality check was done using fastqc (v.0.9.4, Babraham Bioinformatics) and raw reads were aligned to the NCBI37/mm9 version of the mouse genome. Tophat5 (v.2.0.4) was used for alignment with the following parameters: –bowtie1,–max-multihits 5,–genome-read-mismatches 1–segment-mismatches 1–segment-length 20–splice-mismatches 0. Gene expression was quantified with cufflinks (v.2.0.2) with the following parameters: -N, -u. Further, gene expression values were normalized for library size with cuffnorm, with the following parameters: -o -L–library-norm-method classic-fpkm. Cuffdiff was used to find differential gene expression among conditions with the following parameters: -num-threads 16, -multi-read-correct,

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

-frag-bias-correct. Differentially expressed genes were considered with q-value of less than 0.05 (adjusted P values found using an optimized false discovery rate approach).

ChIP–seq data processing. SRR389293 and SRR389294 images were downloaded from the Gene Expression Omnibus website[6]. The quality check was performed using fastqc (v.0.9.4, Babraham Bioinformatics). ChIP–seq reads were aligned to the mouse reference genome NCBI37/mm9 (accessed July 2007) with Burrows–Wheeler Aligner (bwa, v.0.5.9-r16) allowing 0–1 mismatches. Uniquely aligned reads were converted to BED format. Macs14 (v.1.4.1 20110622) was used for peak calling using the following parameters: -t SRR389393, -c SRR389294 -f BED -g mm. Other parameters were used as default. PeakAnalyzer 1.4 was used to annotate NR5A2 binding sites. Motif enrichment was identified with the MEME suite using the default parameters. Reads were directionally extended to 300 bp and, for each base pair in the genome, the number of overlapping sequence reads was determined and averaged over a 10-bp window to create a wig file to visualize the data in the University of California Santa Cruz (UCSC) genome browser.

Principal component analysis. The Pearson correlation was calculated from the expression value (expressed as fragments per kilobase of transcript per million mapped reads) of each gene for each sample by using the 'cor' command in R (https://www.r-project.org/). Principal component analysis was performed using the 'prcomp' command in R, from the correlation value of each sample.

GSEA. The list of genes was ranked by the 't-stat' statistical value from the cuffdiff output file. The list of pre-ranked genes was then analysed with GSEA for Gene Ontology (GO) database. Significantly enriched GO terms were identified using a false discovery rate q value of less than 0.25. The analyses were carried out as defined in http://www.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html?Interpreting_GSEA.

Singular enrichment analysis. The differentially expressed gene sets were computed at the molecular signature dataset of GSEA, using the Biological Processes dataset.

Human transcriptome RNA-seq analysis. Histologically normal fresh frozen pancreatic tissue samples ($n=95$) from patients with pancreatic cancer ($n=79$) (Mayo Clinic; Memorial Sloan Kettering Cancer Center) or from organ donors ($n=16$) (Penn State College of Medicine and Gift of Life Donor Program) were used. Histological review was performed at each participating institution. Subjects of self-reported non-European ancestry and those with a history of neo-adjuvant therapy before surgery were excluded from the study. All relevant ethical regulations were followed. The project was approved by the Institutional Review Boards of Memorial Sloan Kettering Cancer Center, the Mayo Clinic and Georgetown Universities; exemption for the work with human tissue samples at NIH was approved by the NIH Office of Human Subject Research. Human samples were anonymized. Written informed consent was obtained from human subjects.

RNA was isolated with the mirVana kit (Ambion). Samples with RNA integrity numbers of greater than 7.5 (1 μg) were subject to massive parallel paired-end sequencing on the Illumina HiSeq2000 sequencing platform (TruSeq v.3 chemistry) at the CCR Sequencing Facility of the National Cancer Institute, as previously described[27]. Read alignment (MapSplice[28]) and gene expression quantification (in transcripts per million mapped reads, RSEM v.1.2.14[12,29]) was performed using settings and reference data published by The Cancer Genome Atlas University of North Carolina pipeline (https://webshare.bioinf.unc.edu/public/mRNAseq_TCGA/, UNC_mRNAseq_summary.pdf) and the 'UCSC gene' track (hg19/GRCh37) for gene annotation. We compared human samples in the top and bottom quartile of *NR5A2* mRNA levels for expression of genes identified as differentially expressed in the two mouse strains ($\log_2$-fold change >1 or <-1) and a set of random genes; only those genes for which a bona fide orthologue was found in the human genome were included (upregulated, $n=718$; and downregulated, $n=64$). A random list of genes was used for comparison ($n=173$). Logistic regression models based on upper quantile normalized transcripts per million mapped reads expression values were conducted using R and Bioconductor with adjustments for age, gender and study.

Protein expression–quantitative trait locus analysis. Protein expression–quantitative trait locus analysis was conducted using human PDAC tissue samples from patients of European ancestry, by assessing NR5A2 protein levels in formalin-fixed paraffin-embedded tissue samples ($n=110$) on tissue microarrays generated at the Mayo Clinic. Immunohistochemistry was performed with rabbit polyclonal anti-NR5A2 antibodies (HPA005455, Sigma-Aldrich) at 1:300 dilution. NR5A2 staining was scored based on intensity (on a scale from 0–3: 0, negative; 1, weak; 2, positive; and 3, strong) and the proportion of reactive cells (0–100%). Histoscore was calculated as staining intensity × percentage of cells. When more than one core was available from a given tumour, the mean score was used. TaqMan genotyping for the single nucleotide polymorphism marking the PDAC risk locus on chr1q32.1 in *NR5A2* (rs3790844; genotyping assay: C__27483560_10, Thermo Fisher Scientific) was performed on blood DNA from the same patients. Protein expression–quantitative trait loci were assessed by linear regression for additive genetic effects on NR5A2 expression based on mean histoscore and quantiles of mean histoscore (group 1, histoscore of less than 120; group 2, histoscore of 120 to 149; group 3, histoscore of 150 to 169; group 4, histoscore of 170 to 209; and group 5 histoscore of 210 and above) with adjustments for age, gender, body mass index and tissue microarray slide. Comparison of genotypes and NR5A2 expression was performed using mean histoscore quantiles. The Institutional Review Boards of the participating institutions approved the project.

Quantitative RT–PCR. Total RNA was treated with DNase I (Ambion) for 30 min at 37 °C and cDNAs were prepared according to the manufacturer's specifications, using the TaqMan reverse transcription reagents (Applied Biosystems, Roche). qRT–PCR analysis was performed using the SYBR Green PCR master mix and an ABIPRISM 7900HT instrument (Applied Biosystems). The sequence of the primers used is provided in Supplementary Table 4. Expression levels were normalized to endogenous *Hprt* mRNA levels using the $\Delta\Delta C_t$ method. The results shown are representative of at least four biological replicates.

Immunoprecipitation and western blotting. For immunoprecipitation of proteins from fresh total pancreas lysates, a piece of mouse pancreas was isolated and minced in 50 mM Tris–HCl pH 8, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40 containing 3× phosphatase inhibitor cocktail (Sigma-Aldrich) and 3× EDTA-free complete protease inhibitor cocktail (Roche). Lysates were briefly sonicated until the protein solution was clear, cleared for 10 min at 11,000 r.p.m. at 4 °C and then the supernatant was recovered.

Antibody-coated protein A or protein G dynabeads (Life Technology) were used for immunoprecipitation. In brief, beads were washed three times with PBS and incubated with anti-NR5A2 or normal goat IgG overnight at 4 °C. After washing three times with PBS and twice with coupling buffer (27.3 mM sodium tetraborate, 72.7 mM boric acid), the dry beads were incubated overnight at 4 °C in freshly prepared 38 mM dimethyl pimelimidate dihydrochloride in 0.1 M sodium tetraborate. Afterwards, beads were washed three times with coupling buffer and once with 1 M Tris pH 9. Then, 1 ml of the Tris solution was added to the beads and incubated for 10 min at room temperature with rotation to block amino groups and stop crosslinking. Finally, beads were washed three times with storage buffer (6.5 mM sodium tetraborate/boric acid) and stored at 4 °C until used. Protein lysates (1–5 mg, cells; 10–15 mg, tissues) were then incubated overnight at 4 °C with antibody-coated dynabeads (Thermo Fisher Scientific). Bound immune complexes were washed twice with lysis buffer containing NP-40, and then eluted by boiling in 2× Laemmli buffer (10% glycerol, 2% sodium dodecyl sulphate and 0.125 M Tris–HCl pH 6.8) for 5 min.

For co-immunoprecipitation from transfectants, HEK293 cells transfected with the corresponding plasmids were lysed in 50 mM Tris–HCl pH 8, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40 for 30 min at 4 °C. After a brief sonication, cells were cleared by centrifugation for 10 min at 11,000 r.p.m. and the supernatant was recovered. Lysates were pre-cleared with protein A- or protein G-agarose beads and 2 μg of normal rabbit IgG (Sigma-Aldrich) and then immunoprecipitated with anti Flag-M2 affinity gel (A2220), EZview Red anti-HA (E6779) (Sigma-Aldrich) for 2 h at 4 °C. The immune complexes were then pelleted by centrifugation and washed twice in NP-40 lysis buffer for 10 min at 4 °C. After washing, 2× Laemmli buffer was added and a fraction of the material eluted by boiling was loaded onto a SDS–PAGE and processed for western blotting. For immunoprecipitation with anti-NR5A2, antibody (2 μg) was added and incubated overnight.

For western blotting, proteins were extracted from pancreatic tissue, isolated acinar cells or cultured cells using either Laemmli buffer or lysis buffer (50 mM Tris–HCl pH 8, 150 mM NaCl, 5 mM EDTA and 0.5% NP-40) supplemented with protease inhibitor and phosphatase inhibitor cocktails. Protein concentration was measured using the BCA reagent (Biorad), or extrapolated when using Laemmli lysis buffer. Proteins were resolved by either by standard SDS–PAGE or 4–20% TGX pre-cast gels (Biorad) and transferred onto nitrocellulose membranes. The antibodies used are listed in Life Sciences Reporting Summary. Densitometry analysis of digitalised western blotting images was performed using Fiji software (NIH).

ChIP. Pancreas tissue was minced, washed with cold PBS supplemented with 3× protease and phosphatase cocktail inhibitors, and then fixed with 1% formaldehyde for 20 min at room temperature. Glycine was added to a final concentration of 0.125 M for 5 min at room temperature. The fixed tissue was soaked in SDS buffer (50 mM Tris pH 8.1, 100 mM NaCl, 5 mM EDTA and 0.5% SDS) and homogenized using a douncer. The supernatant was collected after centrifugation and the chromatin was sonicated with a Covaris instrument for 40 min (20% duty cycle; 10% intensity; 200 cycle), yielding DNA fragments with a bulk size of 300–500 bp. Samples were centrifuged to pellet cell debris. The amount of chromatin isolated was quantified using nanodrop; an aliquot of this material was used as input for final quantification. Samples (0.5–1 mg of chromatin) were diluted with Triton buffer (100 mM Tris pH 8.6, 0.3% SDS, 1.7% Triton X-100

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

and 5 mM EDTA) to 1 ml and pre-cleared for 2 h with a mix of protein A and G (previously blocked with 5% BSA) at 4 °C. Antibody-coated beads were added: anti-NR5A2 (1 μg), anti-JUN (1 μg), anti-JUNB (1 μg), anti-FOS (5 μg), anti-JUND (2 μg), anti-H3K27ac (1 μg) and anti-H3K27me3 (1 μg). Non-related IgG was used as a control. After incubating for 3 h at 4 °C in a rotating platform, beads were successively washed with 1 ml of mixed micelle buffer (20 mM Tris pH 8.1, 150 mM NaCl, 5 mM EDTA, 5% w/v sucrose, 1% Triton X-100 and 0.2% SDS), buffer 500 (50 mM HEPES at pH 7.5, 0.1% w/v deoxycholic acid, 1% Triton X-100, 500 mM NaCl and 1 mM EDTA), LiCl detergent wash buffer (10 mM Tris at pH 8.0, 0.5% deoxycholic acid, 0.5% NP-40, 250 mM LiCl and 1 mM EDTA) and TE (pH 7.5), and then bound molecules were eluted by incubating overnight in elution buffer (containing 1% SDS and 100 mM $NaHCO_3$) at 65 °C, and treated with proteinase K solution (10 M EDTA, 40 mM Tris–HCl pH 6.5, 40 μg ml^{-1} proteinase K). The eluted DNA was purified by phenol–chloroform extraction. After isolation, pelleted DNA was resuspended in 150 μl of nuclease-free water. Gene occupancy was then analysed by real-time PCR using 1 μl of the eluted DNA diluted in a final volume of 10 μl. The sequence of the primers used for ChIP–qPCR is provided in Supplementary Table 5.

Sequential ChIP. Chromatin from total pancreas was isolated as described above. In brief, after the first round of immunoprecipitation all beads were washed as described for the individual ChIP. Then, 10% of the beads were transferred to a new tube and de-crosslinked. The DNA was then extracted with phenol chloroform to check the efficiency of the first ChIP. For all steps involving immunoprecipitation, low protein-binding tubes were used (Eppendorf, 022431081). For sequential ChIP, the immune complexes in the remaining 90% of beads were eluted by incubating for 30 min at 37 °C in 90 μl of freshly prepared TE and 10 mM DTT, diluted to 2 ml with dilution buffer and incubated with anti-NR5A2 and anti-JUN antibodies overnight at 4 °C. Then, immune complexes were incubated with 40 μl of beads for 2–4 h at 4 °C, washed, transferred to a fresh tube and eluted as described for the individual ChIP.

NR5A2 and NR0B2 knock-down. NR5A2 expression was interfered using Mission shRNA lentiviral constructs purchased from Sigma-Aldrich (TRCN0000025966, targeting CCGGGCAGAAGACTACCTGTACTATCT CGAGATAGTACAGGTAGTCTTCTGCTTTTT (NR5A2 sh1) and TRCN 0000025985, targeting CCGGCCCACAACAGACTGAGAAATTCTCGAGAA TTTCTCAGTCTGTTGTGGGTTTTT (NR5A2 sh2)) previously tested for their efficiency. The knockdown efficiency was analysed by qRT–PCR and immunoblotting. NR0B2 expression was interfered using Mission shRNA lentiviral constructs purchased from Sigma-Aldrich (TRCN0000027118, targeting CCGGCGTCCGACTATTCTGTATGCACTCGAGTGCATACAGAATAGTCGG ACGTTTTT (NR0B2 sh1) and TRCN0000027130, targeting CCGGCA AGGAGTATGCGTACCTGAACTCGAGTTCAGGTACGCATACTCCTTGTT TTT (NR0B2 sh2)). Control cells were transformed using scrambled vector (shNT).

HEK293-FT cells (ATCC) were used to produce lentiviral particles. In brief, cells were allowed to reach 60% of confluence and transfected with 25 μg of shNT, sh1 or sh2 plasmids together with 30 μg of psPAX and 10 μg of pCMV-VSVG helper plasmids using $CaCl_2$ 2 M HBSS, as described earlier. After 12 h, the supernatant was collected and replaced with 5 ml of fresh medium. The supernatant was additionally collected 24 h and 48 h after transfection. The medium was filtered (0.45-μm pore) and added to 266-6 cells (at 50–60% of confluence). After three rounds of infection, the supernatant was removed and replaced with fresh medium. After one day of recovery, puromycin (2 μg ml^{-1}) (Sigma-Aldrich) was added. Two days later, the medium was replaced and cells were collected for protein and RNA analysis after 24 h.

NR0B2 lentiviral overexpression. The production of lentiviral particles and cellular infection were performed as described earlier. The medium from the transfectants was collected 12, 24 and 48 h after transfection. Subsequently, 266-6 cells (obtained from I. Rooman) were infected using polybrene (hexadimethrine bromide, Sigma-Aldrich 107689) (5 μg ml^{-1}). After selection with puromycin, resistant 266-6 cells were collected for RNA and protein analysis.

For combined NR5A2 knockdown and NR0B2 overexpression, cells were first incubated with control shRNA- or NR5A2-targeting lentiviruses. After puromycin selection, resistant 266-6 cells were allowed to recover in fresh medium and incubated with a medium containing lentiviral particles for empty vector or a medium containing lentiviral particles for NR0B2–Flag. Two rounds of infection were performed. After 24 h, the supernatant was removed and fresh medium was added. One day later, 266-6 cells were collected for protein and RNA analyses.

Cloning. *NR5A2* cDNA was amplified and Flag-tagged by PCR using the following primers: GGGAATTCATGGACTACAAGGACGACGATGACAAGTCT GGCTAGTTTGG and GGGGAATTCTTAGGCTCTTTTGGCATGCAGCA. The PCR product was ligated into the pJET vector and then sub-cloned into pLVX-puro. The human *NR0B2* cDNA, tagged with a Flag epitope, was subcloned into the lentiviral vector pLVX-puro. The region encompassing −1960 to −655 of the mouse *Ccl8* promoter was amplified by PCR and subcloned in pGL3-basic vector (Promega). The sequence of the inserts was confirmed by enzymatic digestion and Sanger sequencing.

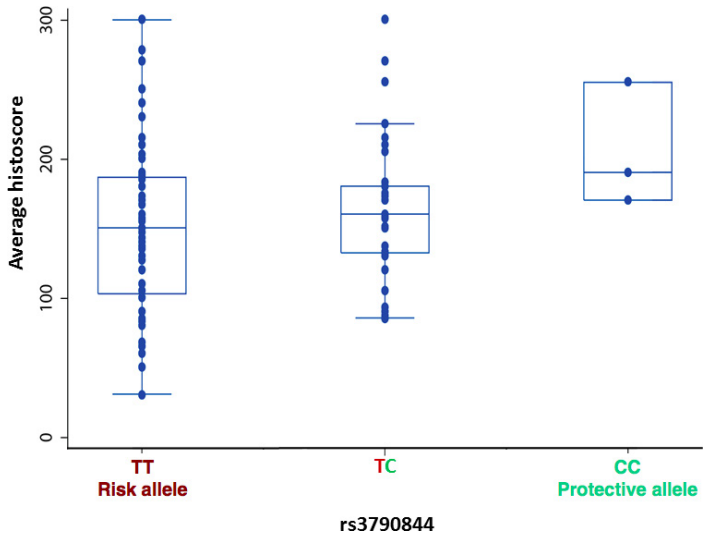
Luciferase reporter assays. HEK293T cells were transfected with plasmids pcDNA3-NR5A2-Flag and JUN-HA, or the corresponding empty plasmids, together with pRL-TK (Promega) constitutively expressing Renilla luciferase for normalization and transfection efficiency control. Plasmids containing the *Ccl8* promoter were included in all the experiments. The amount of transfected plasmids was maintained at a constant level. Transfections were performed using a standard $CaCl_2$ method and measurements were made in triplicate. After 36 h, cells were lysed in passive lysis buffer (Luciferase kit, Promega E2920) and reporter activity was determined using the Dual-Glo Luciferase Assay system (Promega), according to the manufacturer's instructions. Firefly and Renilla luciferase activities were measured using a luminometer.

Other statistical analyses. Comparisons of quantitative data between groups was performed using one-sided Mann–Whitney *U* test in all cases for which there was a prior hypothesis, except for the data shown in Fig. 1g and Extended Data Fig. 5a. Box plots represent the median and first and third quartiles of the data; error bars are generated by R software and represent the highest and lowest data within 1.5× interquartile range. All statistical analyses were performed with Excel or R software.

Data availability. Mouse pancreas RNA-seq data generated in this study have been deposited in GEO with accession number GSE84659. Source Data for Figs. 1–4 and Extended Data Figs 1–10 are provided with the online version of the paper. All other data are available from the corresponding author upon reasonable request.

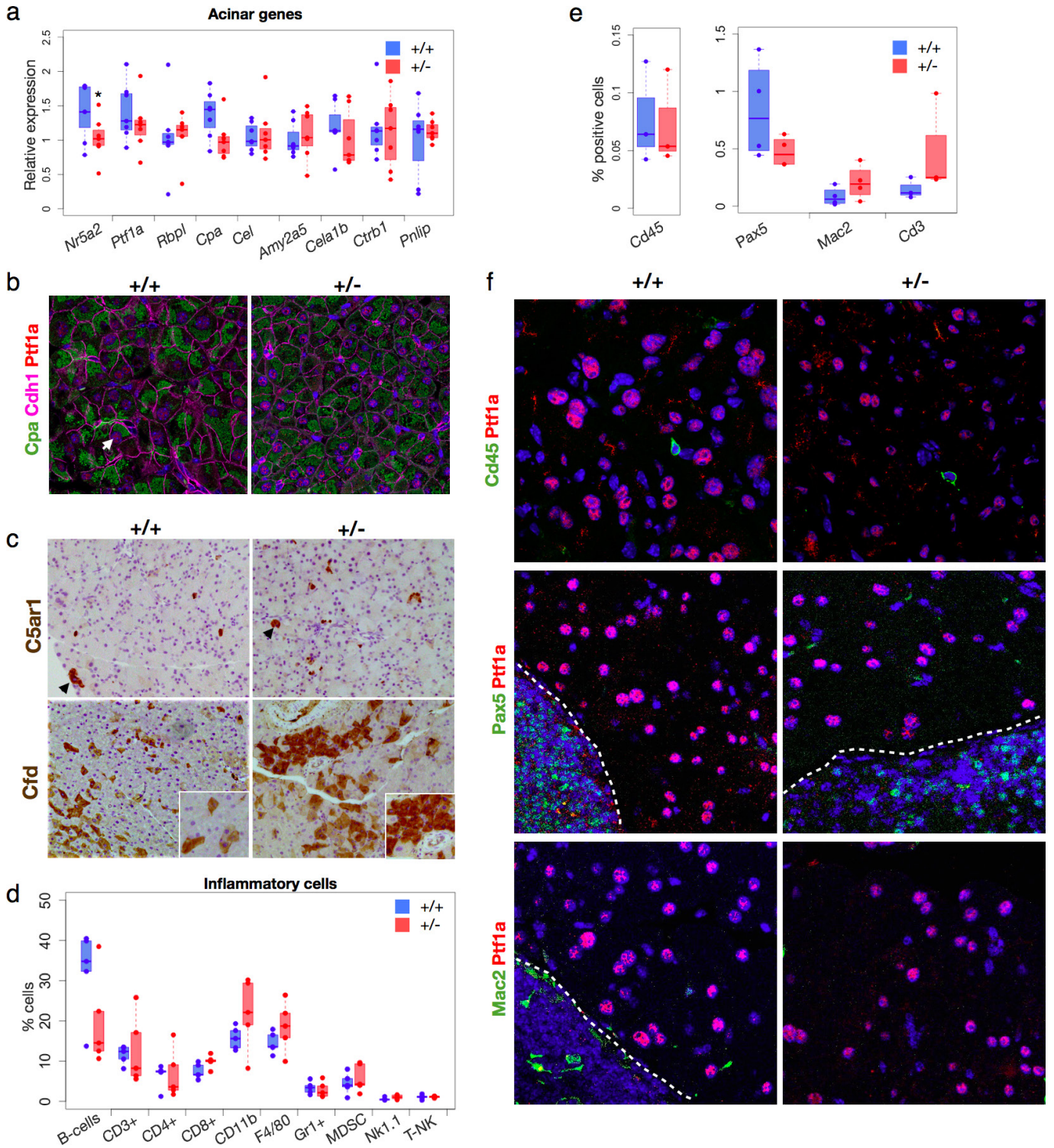
21. Botrugno, O. A. *et al.* Synergy between LRH-1 and β-catenin induces G1 cyclin-mediated cell proliferation. *Mol. Cell* **15,** 499–509 (2004).
22. Behrens, A. *et al.* Impaired postnatal hepatocyte proliferation and liver regeneration in mice lacking c-*jun* in the liver. *EMBO J.* **21,** 1782–1790 (2002).
23. Coste, A. *et al.* LRH-1-mediated glucocorticoid synthesis in enterocytes protects against inflammatory bowel disease. *Proc. Natl Acad. Sci. USA* **104,** 13098–13103 (2007).
24. Clausen, B. E., Burkhardt, C., Reith, W., Renkawitz, R. & Förster, I. Conditional gene targeting in macrophages and granulocytes using LysMcre mice. *Transgenic Res.* **8,** 265–277 (1999).
25. Kawaguchi, Y. *et al.* The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. *Nat. Genet.* **32,** 128–134 (2002).
26. Cendrowski, J. *et al.* Mnk1 is a novel acinar cell-specific kinase required for exocrine pancreatic secretion and response to pancreatitis in mice. *Gut* **64,** 937–947 (2015).
27. Hoskins, J. W. *et al.* Transcriptome analysis of pancreatic cancer reveals a tumor suppressor function for HNF1A. *Carcinogenesis* **35,** 2670–2678 (2014).
28. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
29. Wang, K. *et al.* MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* **38,** e178 (2010).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



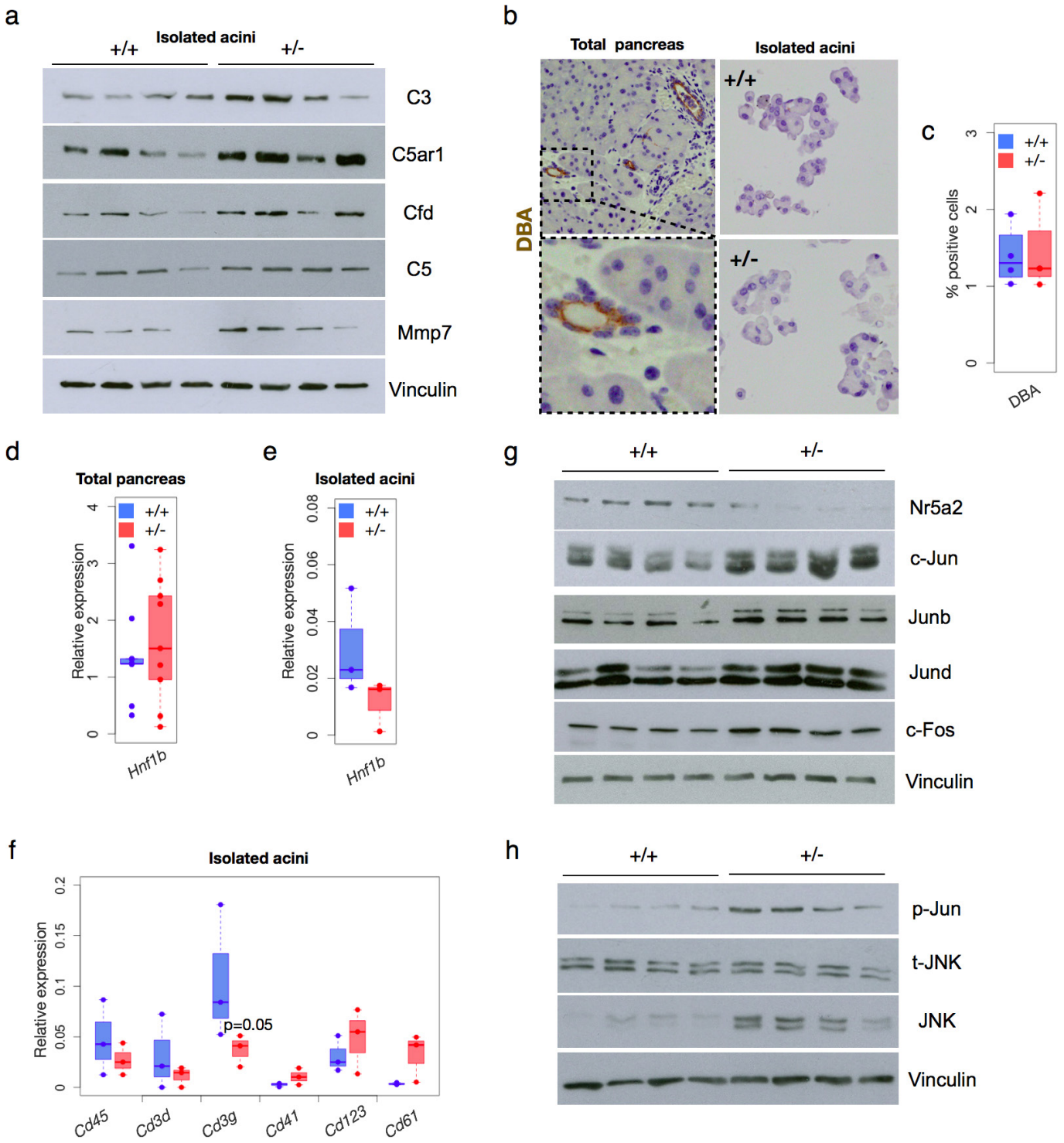
Extended Data Figure 1 | Protein eQTL analysis in human PDAC. Pancreatic tumours ($n = 110$) from patients carrying the risk-increasing allele (T) at rs3790844 express lower levels of NR5A2 protein than those carrying the protective allele (C). NR5A2 expression was assessed using immunohistochemistry, and scored based on percentage of reactive cells and intensity of staining. The analysis was performed for mean histoscore ($P = 0.097$, $\beta = -18.0$; two-sided Wilcoxon test) and mean histoscore quantiles ($P = 0.028$, $\beta = -0.57$; two-sided Wilcoxon test).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



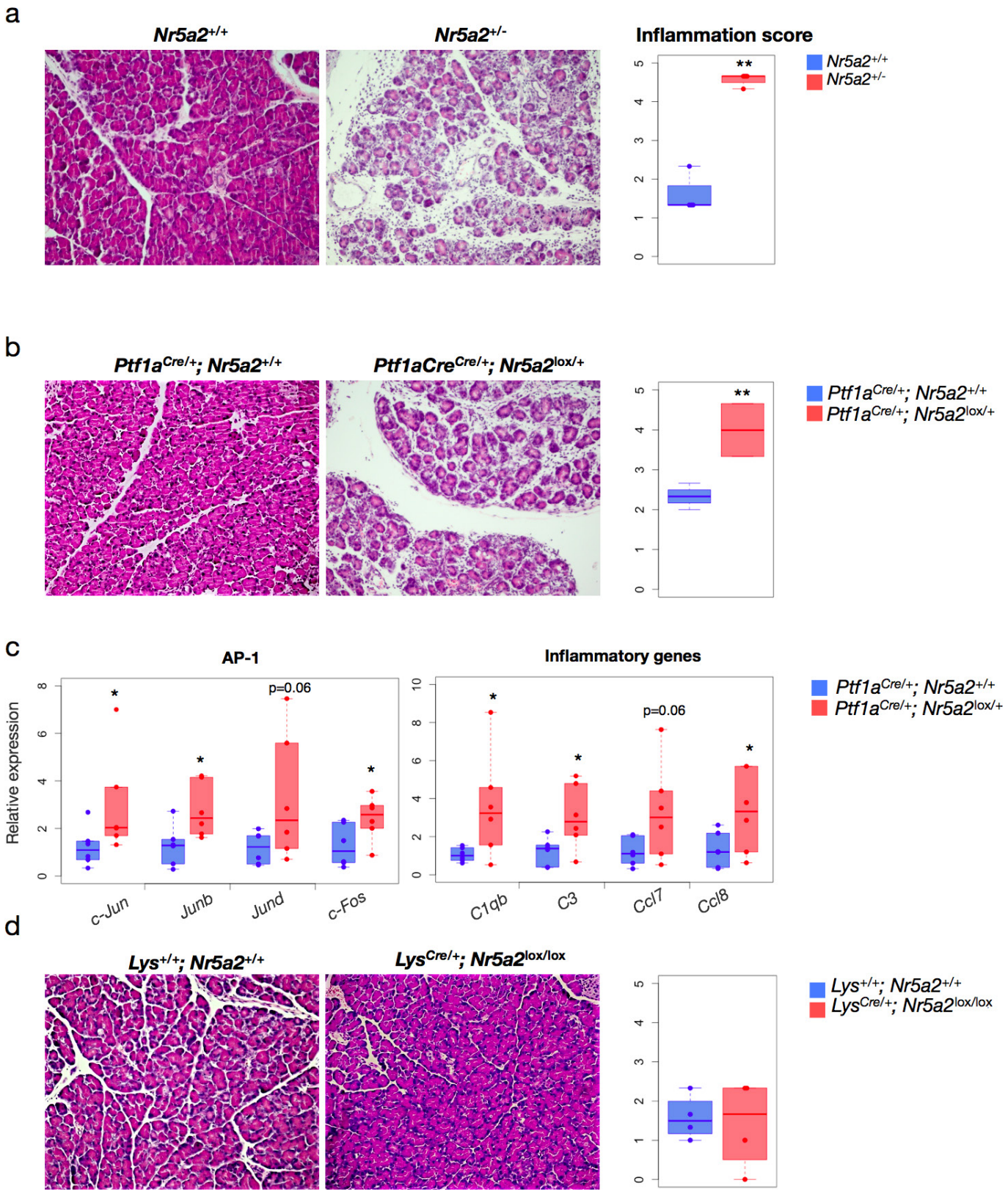
Extended Data Figure 2 | The pancreas of *Nr5a2*$^{+/-}$ mice is histologically normal but displays increased expression of inflammatory genes. **a**, qRT–PCR analysis of the expression of transcripts coding for acinar-related genes in wild-type and *Nr5a2*$^{+/-}$ mice ($n = 7$ per group). Data were obtained from a series of mice that was independent of the series used for RNA-seq. **b**, Immunofluorescence analysis of PTF1a, CDH1 and CPA in the pancreas of wild-type and *Nr5a2*$^{+/-}$ mice ($n = 3$ per group). Arrow, acinus. **c**, Immunohistochemical analysis of expression of C5AR1 and CFD in the pancreas of wild-type and *Nr5a2*$^{+/-}$ mice shows patchy expression in acinar cells (arrows) ($n = 5$ per group). **d**, Percentage of inflammatory cell subtypes (from total cells) in wild-type and *Nr5a2*$^{+/-}$ pancreases ($n \geq 4$ per group) analysed by flow cytometry (two different experiments). **e**, **f**, Quantification of periacinar Cd45^{+} cells in the pancreas of wild-type and *Nr5a2*$^{+/-}$ mice using immunofluorescence on frozen sections. Broken line delineates a pancreatic lymph node, used as a control. Two independent assessments were performed. **e**, Quantification of cells expressing PAX5, MAC2 and CD3 in the pancreas of wild-type and *Nr5a2*$^{+/-}$ mice using immunofluorescence ($n \geq 3$ per group). In **a**, **d** and **e**, one-sided Mann–Whitney *U* test; $^{*}P < 0.05$, $^{**}P < 0.01$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



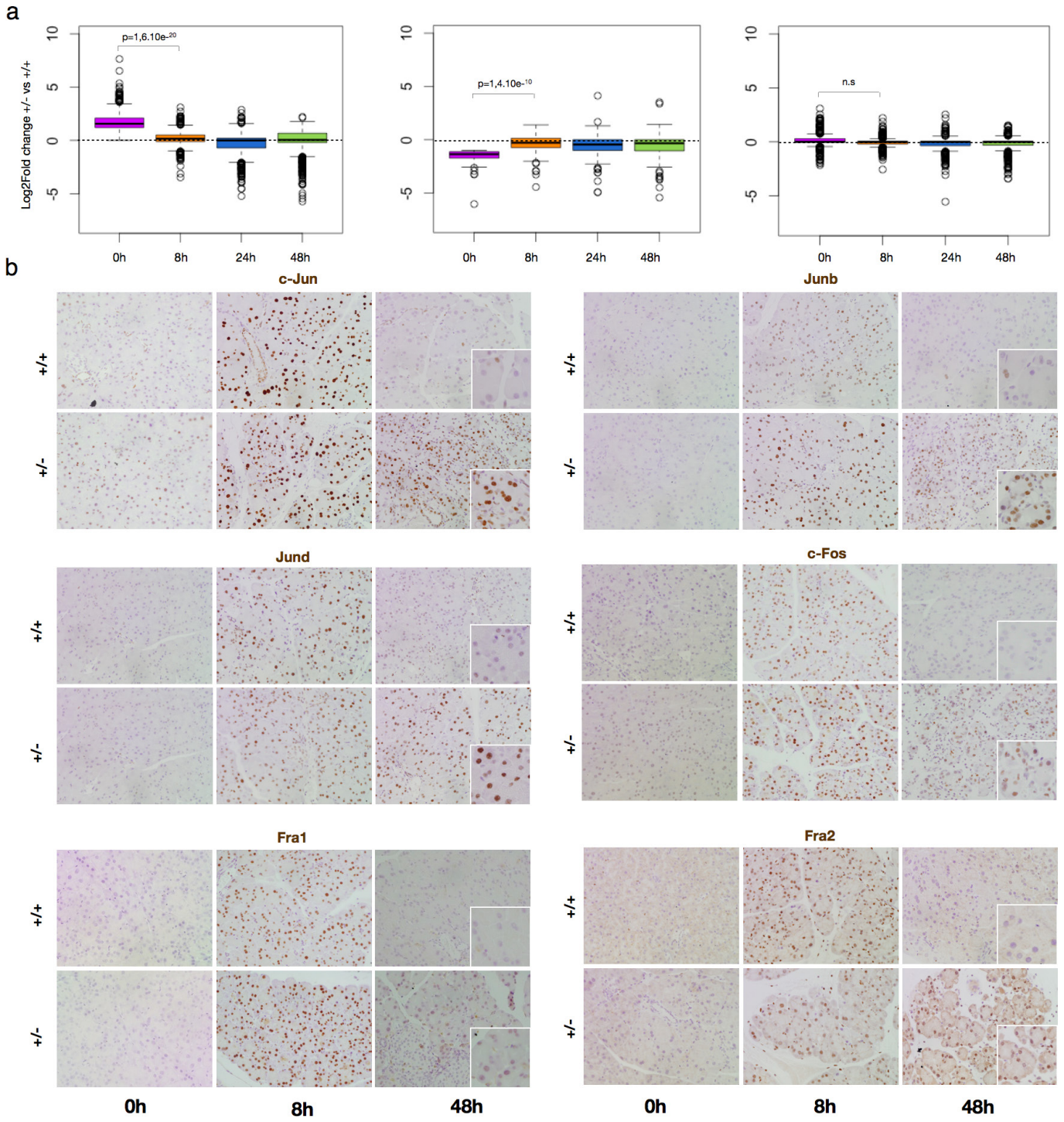
Extended Data Figure 3 | The upregulation of inflammatory markers, AP-1 components, p-JUN and p-JNK in *Nr5a2*$^{+/-}$ pancreases is epithelial-cell-autonomous, as shown by the analysis of isolated primary acinar cells. a, Expression of inflammatory proteins in primary acinar cells from wild-type and *Nr5a2*$^{+/-}$ mice shown using western blotting ($n = 4$ per group). **b**–**f**, Primary acinar cell fractions from wild-type and *Nr5a2*$^{+/-}$ mice largely depleted of DBA^{+} ductal cells (**b**, **c**), show reduced expression of the ductal cell marker HNF1β (**d**, **e**) and inflammatory cell markers (**f**) compared to total pancreas ($n = 4$ per group). Inset in **b**, DBA-labelled duct. Two independent experiments were performed. **g**, **h**, Expression of AP-1 components and JNK in primary acinar cells from wild type and *Nr5a2*$^{+/-}$ mice using western blotting. NR5A2 is expressed at reduced levels in *Nr5a2*$^{+/-}$ pancreases ($n = 4$/group). In **c**–**f**, one-sided Mann–Whitney *U* test; $^{*}P < 0.05$, $^{**}P < 0.01$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



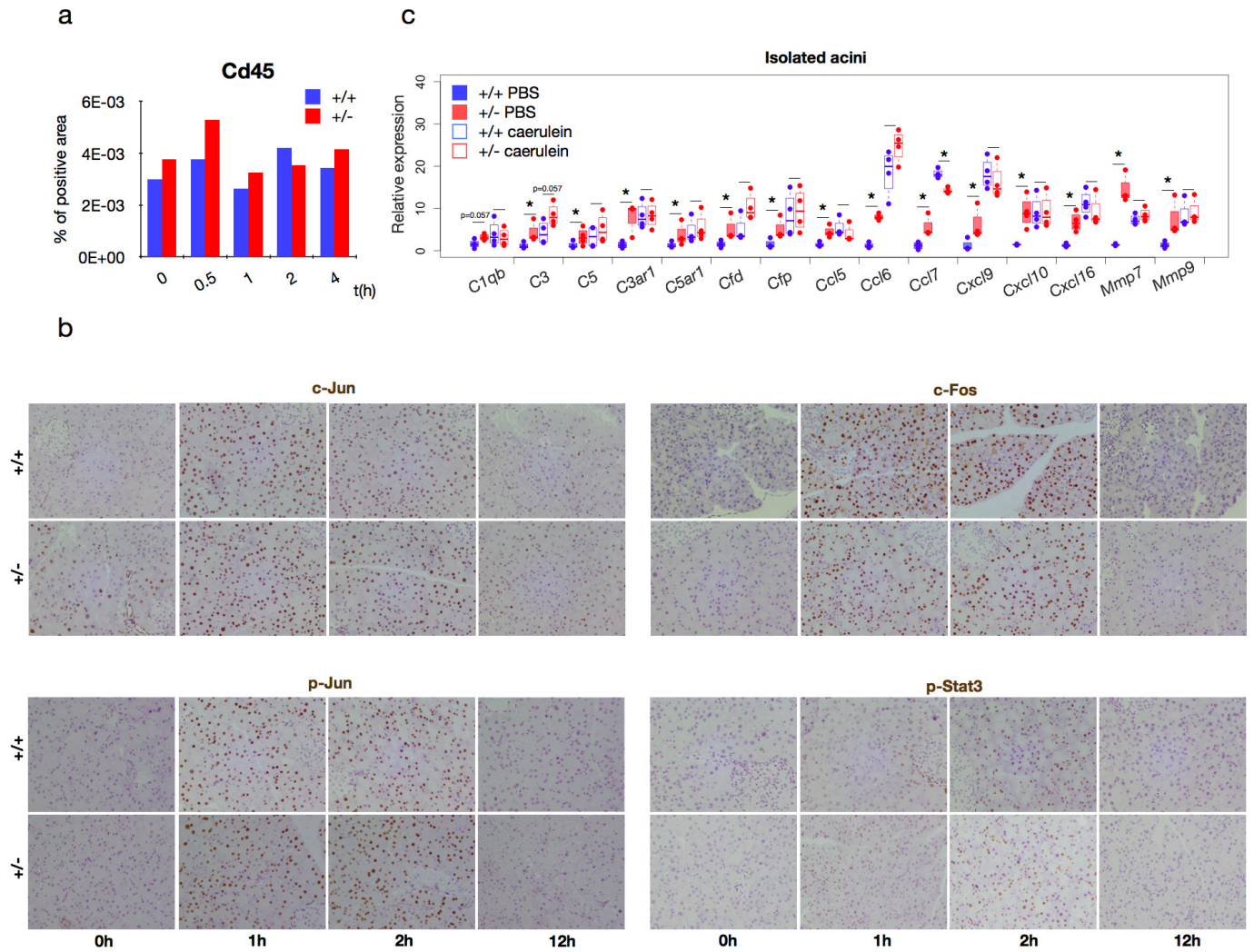
Extended Data Figure 4 | The defective pancreatic response to damage is epithelial-cell-autonomous. a, Constitutive *Nr5a2*$^{+/-}$ mice display more severe pancreatitis upon administration of seven doses of caerulein (given once per hour). **b**, **d**, This severe phenotype is recapitulated at 48 h in mice harbouring a heterozygous deletion of *Nr5a2* in pancreatic epithelial cells (**b**) but not in mice in which both alleles of *Nr5a2* are inactivated in myeloid cells by Cre activation from the lysozyme endogenous locus (*Lys*) (**d**). This experiment was performed once for the conditional mice; for *Nr5a2*$^{+/-}$ mice, more than four independent experiments were performed. Representative histological images are shown. Semi-quantitative inflammation scores corresponding to the experiments are shown in **a**–**d** ($n \geq 4$ per group). **c**, qRT–PCR analysis of the expression of transcripts coding for AP-1 and inflammatory genes in control and *Ptf1a*^{cre};*Nr5a2*$^{lox/+}$ mice ($n = 6$ per group). In **a**–**d**, one-sided Mann–Whitney *U* test; $^{*}P < 0.05$, $^{**}P < 0.01$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 5 | *Nr5a2* haploinsufficiency causes a basal pre-inflammatory state similar to that associated with the early stages of pancreatitis. a, Comparative expression (wild-type versus *Nr5a2*$^{+/-}$ mice) of the upregulated (left), downregulated (middle) or control (right) genes over time, after induction of pancreatitis. RNA-seq analysis was performed once. One-sided Student's *t*-test. **b**, Immunohistochemical analysis shows persistent overexpression of AP-1 components during the recovery period after induction of acute pancreatitis (one dose per hour for seven hours). Representative results of one of five pancreases analysed are shown.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 6 | A single dose of caerulein does not cause inflammation but does induce an upregulation of AP-1 and p-JUN that precedes STAT3 phosphorylation both in wild-type and *Nr5a2*$^{+/-}$ mice. **a**, Quantification of infiltration by Cd45^{+} cells in the pancreas of wild-type and *Nr5a2*$^{+/-}$ mice after administration of one dose of caerulein ($n = 1$). **b**, Immunohistochemical analysis of expression of JUN, FOS, p-JUN and phospho-STAT3 (p-STAT3) in wild-type and *Nr5a2*$^{+/-}$ mice at various time points after caerulein administration ($n = 4$ per group). **c**, qRT–PCR analysis of expression of a panel of inflammatory genes in isolated acini treated with PBS or caerulein (100 pM). Data are shown relative to values of wild-type acini incubated with PBS ($n = 4$ per group). Two independent experiments were performed. In **c**, one-sided Mann–Whitney *U* test; $^{*}P < 0.05$, $^{**}P < 0.01$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

a

***C1qb* promoter (-1581;-655)**

GGTCCTGATGTATAGACCCTGTCTGCTCCAAGTCTGTCCCCTCTTCATATACATAATACTGTGTCT
GAGGGCCCTGGACTTTGAAGGTTCCTCCCCTGTTTCACCAACAGCCCCAGGAAACAGGCTCTA
CCATCTCTGATGTAGACCTGAGAGAACCCCAGTTAAGAACATGCTTTGGCCTATATGACTGGAAA
CTGGGGTTCAAACTTAGGGAGTTCAGATTTTCTTACTGGCAATTTACTCAGGAATCAGAGACTCA
GCCTAGGACCCTCAGAGGGACCCGAGGCTGGTGATCACGATGCACACCTGTAATTCCAGTACG
TGGGAGACAGAGAGACCACGGGATTTTAAGTCTGAGGCCAACCTGGGATAAAAAGAGTGAGAC
CCTGCCTCAAAGCTAAGTAAGTGACAGCCGGTTCCCTGGGCACATGGCCCTCAGCTCTCCCAT
CTCTTTTGAACCTGGCTGAGTCATCCTCTTTTCCTCTTGTCCCTCTGGGTATTGGGTGACACCTG
GCTGTACTTACACACAGTCACAGCAGCACAGCTAGACCCTGGCCCCAGAGCTCTCTGGCTTCT
GGGTCTGCTCTGGAGCATGGGTTCTAAAATTCACTGATTTTTCTGTGACATTTACCACAGAAACAT
TTAATGTTTAACTCCAGCCTCCCATAAAGGGATGCCCTCAAAGGACAGGACTCTTTGCTACCCCT
TAAGGCCTTCTTCCCTAGACTATAGCTCATAGTATTTGTGCCCCCCTCCAACAAAGTCCGGTTGG
ATGATAACTCATGCCTCTGGACCTCATTTCCCTGAGTGTAAAGTGAGGGACAGGGTGGGCACAG
TGACCTCCAAGCCCCTCCGCAGCTCTGTCTTGCTGTCATTCCATAGGGAGCAGTCCCTTCTCTG
TGGCTGTTATTGTCAAGGGCTCCAAGC

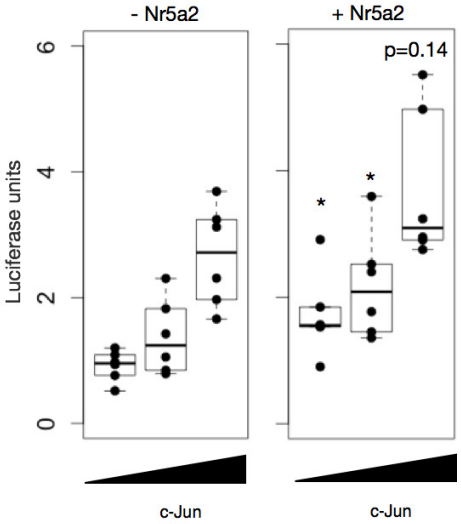
***Ccl7* promoter (-590;+364)**

AAAGAATGCTTTGGGATGGCCCTCAACTTCAGCCGTCTACCTACAACCAAGTACTGGCCAGCTG
ATGCCAGTCAACTTTCCCAAATAAAGAGCTGTTTTGTTTTTAGTAATTAAAAATTATATTAGACACAG
GGTTTGGAAATGTTTTATGATTTTTGTTTGTCTGTTTGCAAACATCTTTTCCTCTGTGGTTTGAAAA
GTTAGCAGGACTTTTCATCTTCAAGTGCATTTTTAAATTCAGTCAAAATTTTAGTGCCTGTTTCTAG
ACTCCAAGCTCCTTGGGATGTTGCTCGAAGGGGGAAATAGGAATCCCTTGACTCTGCCATCTCT
GCTTCCTTCCTCATTCTACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCTTTTCTTTCCAATA
GTAGCTGAGCTGGTTGGAATAGCCAGAGTTCACAGTACAGTGTTTTATCTGGACAGAGTCACCA
CCTCATTTTCCATTCCTGGTGTTAAATGGAAATAGGCCAGCCTCACATTACACTCTCCTGTGATTC
ATTCCAAAAGAAGATCCCTCCCTATCGGCTGCCTATAAAAGAGGAGGCAGAGCTGCCAGAAGAG
CAGAGAAGCAAGGCCAGCACAGAGTCTGCCAGCTCTCACTGAAGCCAGCTCTCTCACTCTCTT
TCTCCACCATGAGGATCTCTGCCACGCTTCTGTGCCTGCTGCTCATAGCCGCTGCTTTCAGCAT
CCAAGTGTGGGCCCAACCAGGTGAGACTCACCCTGCCTCCTGTTTCCCAGAGCATGCTGTCCC
TTCTCTGGGACATTTCATAGAAAATTACAGAACAAGGTTTACCCATAGTCTCACATTTGCTGGCTTT
GCCAAAGTAGGACAATTAAAAGAAGGACAAGAAGATACTTATAGTCCAATGATAATTCTAGAACTTG
AACCTGATTCCCTTAGGACTCATTTGTGGAAACAAGATAG

***Ccl8* promoter (-1960;-655)**

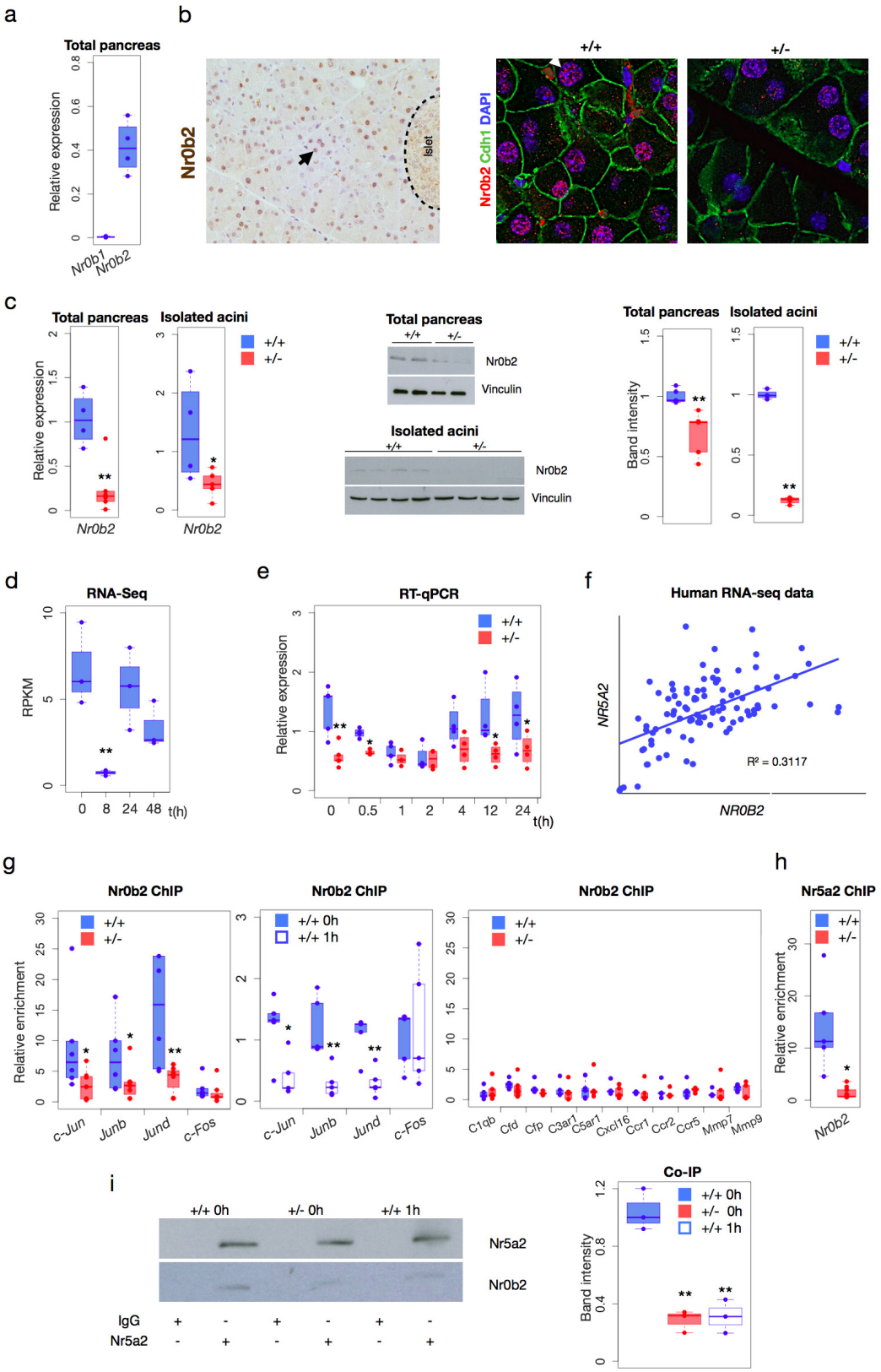
GGAAGAACCCTCCACACTGAGCTCTTCCAAATTTACTTTGTCCGGAAGAACCCTCCACACTGAT
TTCCATAGTAAAAAATTCTGTAGTAGGTGGCAGAGTATGTGTGCAGACT
TCTATTTGGGTCCTCATTTTTATACATCAATCAGTGTGTCTGGTTTTGTGCTCTAACTATTAAGATTAT
AAACTCTTAGATTCATTTAATACTAAAGGTGATAAACACTTTGGTTCAAACATTTCACCTCAAACTTG
TACCAACTCCTGCTGCCTTTGTGGCCTGCAGAAATGCTGACTCGTGTTCTTCTTCACTTACGACG
ACAGTTGGGTATTGTAAGTGCCTAAGCTTGAGTCCAACAACATTCCCAATTCTCCGTTTCAATGTT
TCTTAAATCATTTCTTTTCATGTTTAGATATTTTGAAAACTGTCTTAGGCATTTGAAATTCTCTTTAATA
TTTTCTAAGTGACAAACAGTGTGATAAATCATGTCTGCAGGGAAGGTTCAAATTAGTAAGCCATGC
CCTTAGTCCATCATGCTTTCTTATGTCTGTTTGCCTTCTCACACTTCAAACCTTGTCTGAAACAAG
TCTTCATTTCGGAAAAAGACTCTGTAGACACAGGGAAAATAATCACTTGGATTAAGGAGTGTGATG
ACTCATGACATACTGAGGAGTTTGGCCTTGCTTGCTTTGGTTTCTTTGCGTTGTTTTGTTGTTGT
TGGGGTTTTTTGTGTTTGTTTGTTTGTTTGTTTGTTTGTTTTAGGTCATGGTGTCTTCTTGCTT
GTTATAACTTTAGATATAATGTACTACTGTGTGTTGCATTTAAATGTTGTAGCTGACAGAGGTCAAAA
CCCAGAAAAATCATAAAATATTAGCCATTAGCTCATGGTGGCTTAATAGATTGGATATAATTCAGAGA
TTTAATTCTGAGTGGCTTTACTAACTATATCCCTCACTTGAGGATGTTCTTCCTGCCTACTAGTCAG
GAATCTTTGCTTACACTGACGAGGAATGAGACTGAGCATAACACTATTTGGCCTCCAAACACATT
GAACAGGATGCTTGGCCAACAGAGTTATTGTGATTGAAGTTCCCAGACAAAAGTCATACATACATT
AAGGATTCTGCAAGAAAACAGCTCTGTGATGTAGAGAGCTATGCTCCTGCAGGATGCCACCCGC
TGCAGGATCCCTGCCCCTCCAGAATCCACACGTGGTGCCCAGTTCTCCAGGACTTTCAGGCTA
GGCCAGACATAGTAGAGAAACTCTGTTATTACAGCATTTTTCAGACTGATATGGGAGAATTATACAA
CAAGGACGGCCTGGGGACAAAGTAAATTTGGAAGAGC

b



Extended Data Figure 7 | NR5A2 cooperates with AP-1 to regulate inflammatory gene expression. a, Analysis of putative NR5A2 and AP-1 binding sites in the proximal promoter of *C1qb*, *Ccc7* and *Ccl8* using the JASPAR algorithm (http://jaspar.genereg.net/search?q=&collection=CORE&tax_group=vertebrates). Sequence matrices for NR5A2, FOS:JUN, FOS, JUN, JUN (var.2), FOSL1, FOSL2 and BATF:JUN were computed. Motifs with a score of greater than 7.5 for NR5A2 (blue) and AP-1 (orange) are highlighted. Additionally, a manual search for the NR5A2 binding motif CAAGGNCA was performed. Purple, genomic regions amplified in the sequential ChIP–qPCR experiments shown in Fig. 3f. The sequences of the 400 nucleotides upstream and downstream of each amplicon are also shown. **b**, *Ccl8* luciferase promoter–reporter activity (−1960 to −655) using HEK293 cells and increasing concentrations of JUN-coding plasmid in the absence (left) or presence (right) of NR5A2. Data shown corresponds to the mean of six independent experiments. In **b**, one-sided Mann–Whitney *U* test; $*P < 0.05$, $**P < 0.01$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

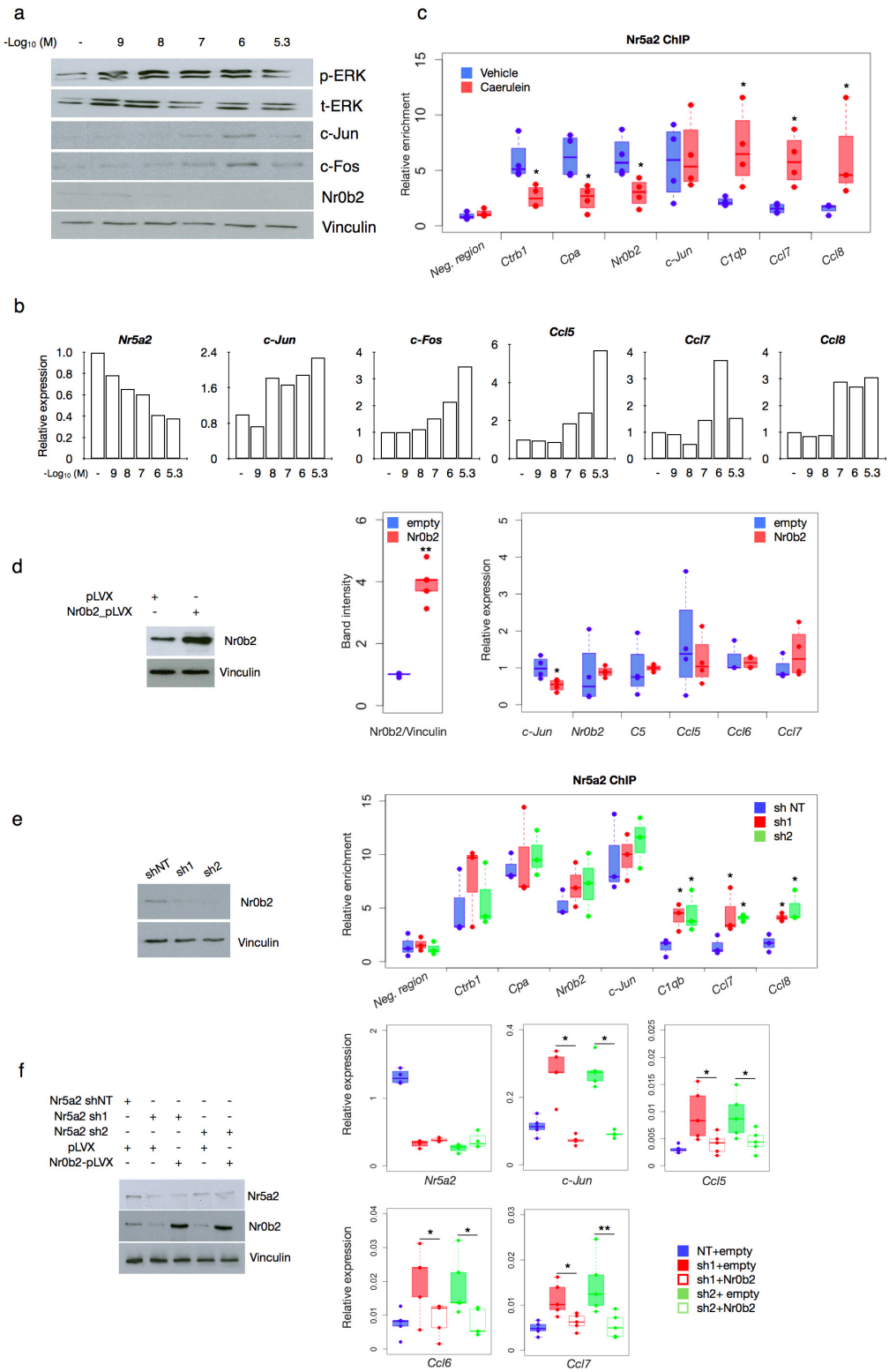


Extended Data Figure 8 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

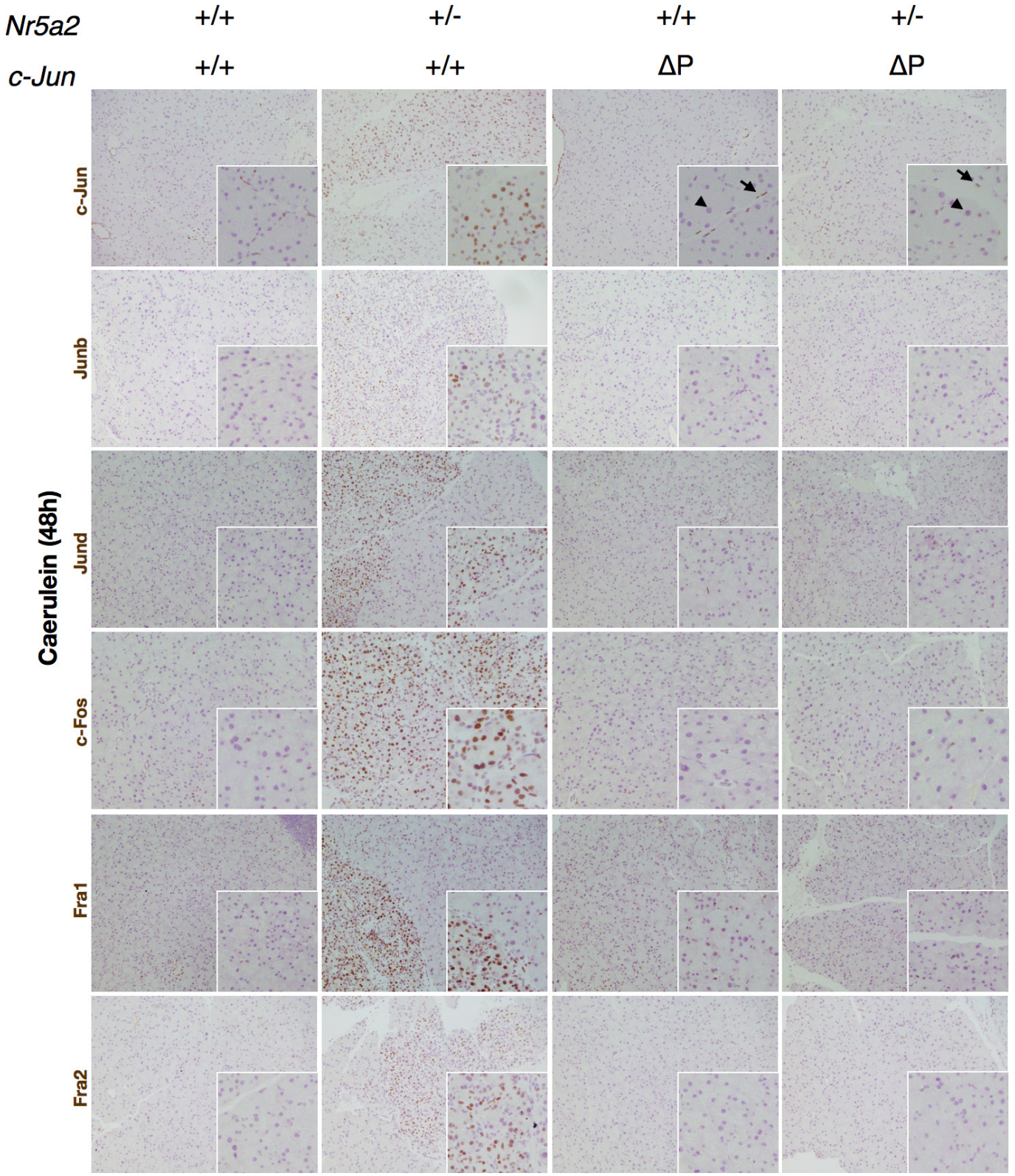
Extended Data Figure 8 | NR5A2 regulates AP-1 expression, in part through the modulation of NR0B2 and its recruitment to AP-1 gene promoters. **a**, Expression of *Nr0b1* and *Nr0b2* transcripts in total pancreas and isolated acini of wild-type and $Nr5a2^{+/-}$ mice ($n = 4$ per group). Arrow, acinus; broken line delineates an islet. **b**, Immunohistochemical and double immunofluorescence analysis showing acinar distribution of NR0B2 in wild-type pancreas, and reduced expression in $Nr5a2^{+/-}$ pancreases. Acinar cells are delineated with anti-CDH1 antibodies ($n = 5$ per group). **c**, Reduced expression of *Nr0b2* mRNA and corresponding protein in total pancreas and isolated acini of wild-type and $Nr5a2^{+/-}$ mice. Densitometric quantification of NR0B2 expression relative to vinculin ($n \geq 4$ per group). **d**, **e**, Expression of *Nr0b2* mRNA in wild-type mice on induction of mild acute pancreatitis (**d**) ($n = 3$ per group) or on administration of a single dose of caerulein (**e**) ($n \geq 3$ per group). **f**, Correlation of *NR5A2* and *NR0B2* mRNA expression in normal human pancreas using RNA-seq. **g**, ChIP–qPCR analysis of the occupancy by NR0B2 at the AP-1 (left) and inflammatory gene promoters (right panel) in wild-type and $Nr5a2^{+/-}$ mice. In the left and right panels, data are shown relative to control IgG and an unrelated genomic region ($n = 3$ per group). ChIP–qPCR analysis of NR0B2 on the promoter of AP-1 genes shows reduced occupancy in wild-type mice 1 h after administration of one dose of caerulein. Results in the middle panel are normalized to enrichment in wild-type mice ($n \geq 6$ per group). **h**, ChIP–qPCR analysis of the occupancy of the *Nr0b2* promoter by NR5A2 in wild-type and $Nr5a2^{+/-}$ mice. Data are shown relative to control IgG and an unrelated genomic region ($n \geq 5$ per group). **i**, Co-immunoprecipitation of NR5A2 and NR0B2 in wild-type and $Nr5a2^{+/-}$ pancreases under basal conditions or 1 h after administration of a single dose of caerulein. Densitometric quantification of NR0B2 bands (right) ($n = 3$ per group). At least two independent experiments were performed. In **a**–**i**, one-sided Mann–Whitney *U* test; $*P < 0.05$, $**P < 0.01$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 9 | NR0B2 has an important role in the dynamic regulation of inflammatory genes by NR5A2. a–**c**, Validation of 266-6 cells as a model for mechanistic studies. **a**, Dose-dependent effects of caerulein on ERK activation, AP-1 expression and NR0B2 expression shown using western blotting. **b**, qRT–PCR analysis showing caerulein-induced changes in expression of *Nr5a2*, *AP-1* and inflammatory genes. **c**, ChIP–qPCR analysis showing changes in NR5A2 occupancy of the promoters of acinar (*Ctrb1*, *Cpa* and *Nr0b2*), *Jun* and inflammatory genes (*C1qb*, *Ccl7* and *Ccl8*) 30 min after treatment with caerulein (4 independent experiments) These findings largely recapitulate the observations made in the mouse pancreas. **d**, Forced overexpression of NR0B2 leads to reduced expression of *Jun* mRNA but does not affect expression of inflammatory genes (four independent experiments). **e**, Effects of NR0B2 knockdown on NR5A2 binding to the promoter of acinar, *Jun* and inflammatory genes (three independent experiments). **f**, Combined NR5A2 knockdown and NR0B2 overexpression showing that higher levels of NR0B2 are associated with reduced expression of inflammatory gene transcripts, a situation that mimics normal pancreas under basal conditions in wild-type mice (four independent experiments). At least two independent experiments were performed. In **c**–**f**, one-sided Mann–Whitney *U* test; $*P < 0.05$, $**P < 0.01$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | JUN is required for the overactivation of AP-1 that is observed in *Nr5a2*$^{+/-}$ mice during caerulein-mediated pancreatitis. Immunohistochemical analysis of the expression of JUN, JUNB, JUND, FOS, FOSL1 and FOSL2 in the pancreas of control (*Nr5a2*$^{+/+}$, *Nr5a2*$^{+/-}$ and *Nr5a2*$^{+/+}$;*JunΔP*) and *Nr5a2*$^{+/-}$;*JunΔP* mice 48 h after the initiation of pancreatitis ($n = 4$ per group). One experiment was performed. Arrowhead, acinus; arrow, mesothelial cell.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature25492

TGFβ drives immune evasion in genetically reconstituted colon cancer metastasis

Daniele V. F. Tauriello[1,2], Sergio Palomo-Ponce[1,2], Diana Stork[1], Antonio Berenguer-Llergo[1], Jordi Badia-Ramentol[1], Mar Iglesias[2,3,4,5], Marta Sevillano[1,2], Sales Ibiza[1], Adrià Cañellas[1], Xavier Hernando-Momblona[1,2], Daniel Byrom[1], Joan A. Matarin[1], Alexandre Calon[1]†, Elisa I. Rivas[1]†, Angel R. Nebreda[1,6], Antoni Riera[1,7], Camille Stephan-Otto Attolini[1] & Eduard Batlle[1,2,6]

Most patients with colorectal cancer die as a result of the disease spreading to other organs. However, no prevalent mutations have been associated with metastatic colorectal cancers[1,2]. Instead, particular features of the tumour microenvironment, such as lack of T-cell infiltration[3], low type 1 T-helper cell (T_H1) activity and reduced immune cytotoxicity[2] or increased TGFβ levels[4] predict adverse outcomes in patients with colorectal cancer. Here we analyse the interplay between genetic alterations and the tumour microenvironment by crossing mice bearing conditional alleles of four main colorectal cancer mutations in intestinal stem cells. Quadruple-mutant mice developed metastatic intestinal tumours that display key hallmarks of human microsatellite-stable colorectal cancers, including low mutational burden[5], T-cell exclusion[3] and TGFβ-activated stroma[4,6,7]. Inhibition of the PD-1–PD-L1 immune checkpoint provoked a limited response in this model system. By contrast, inhibition of TGFβ unleashed a potent and enduring cytotoxic T-cell response against tumour cells that prevented metastasis. In mice with progressive liver metastatic disease, blockade of TGFβ signalling rendered tumours susceptible to anti-PD-1–PD-L1 therapy. Our data show that increased TGFβ in the tumour microenvironment represents a primary mechanism of immune evasion that promotes T-cell exclusion and blocks acquisition of the T_H1-effector phenotype. Immunotherapies directed against TGFβ signalling may therefore have broad applications in treating patients with advanced colorectal cancer.

Progression of colorectal cancer (CRC) generally coincides with successive alterations in four signalling pathways: WNT, EGFR, p53 and TGFβ[5,8]. Mice bearing compound mutations in these four pathways were recently shown to enable the study of CRC metastasis[9–11]. We crossed mice bearing conditional alleles in homologues of four key human CRC mutations: $Apc^{fl/fl}$, $Kras^{LSL\text{-}G12D}$, $Tgfbr2^{fl/fl}$ and $Trp53^{fl/fl}$ (designated A, K, T and P, respectively)[12–15], and targeted gene recombination to intestinal stem cells (ISCs) by means of the $Lgr5^{eGFP\text{-}creERT2}$ driver[16], which we designated L. We generated eight mouse strains bearing combinations of these mutations (Fig. 1a). Histopathological scoring demonstrated a stepwise increase in prevalence and severity of invasive adenocarcinomas along the linear progression sequence (Fig. 1b, c and Extended Data Fig. 1c–i). Ninety per cent of LAKTP mice developed carcinomas, more than half of which breached all intestinal layers (Fig. 1c). These cancers displayed a histology similar to human tumours, with mostly medium to high degrees of differentiation and abundant desmoplastic reaction (Extended Data Fig. 1d–j). Forty per cent of LAKTP mice developed metastases in the liver or lungs, or as carcinomatosis (Fig. 1d and Extended Data Fig. 1k–n), with a median latency of 66 days. Notably, mice bearing triple-mutant genotypes (LAKT, LATP or LAKP) presented with similarly invasive cancers but not metastasis (Fig. 1c). In LAKTP mice, T cells extensively infiltrated the stroma of normal mucosa and adenomas but were largely excluded from adjacent invasive cancers (Fig. 1e and Extended Data Fig. 2a). This exclusion phenotype intensified along the CRC mutational sequence (Fig. 1e). Invasive margins of compound-mutant cancers displayed high levels of stromal TGFβ activity, as indicated by the presence of phosphorylated SMAD3 (pSMAD3) (Fig. 1f and Extended Data Fig. 2b) and expression of CALD1 and IGFBP7 (Extended Data Fig. 2c, d)—two TGFβ-induced genes expressed in cancer-associated fibroblasts (CAFs) that predict poor prognosis[4]. These genetic CRC models therefore reproduce key features of the tumour microenvironment (TME) in advanced human CRCs.

We collected fresh samples from multiple primary mouse tumours or metastases and established a mouse tumour organoid (MTO) biobank (Fig. 1g). Quadruple-mutant LAKTP MTOs expanded *in vitro* independently of factors that stimulate WNT, EGF and TGFβ pathways (Fig. 1h and Extended Data Fig. 3). When injected into the caecum wall of syngeneic C57BL/6J recipients, these MTOs engrafted with a success rate of 10 out of 32 (31%) and progressed to fully invasive T3–T4 tumours (Extended Data Fig. 4a, b). Forty per cent of engrafted MTOs produced overt liver metastatic disease. We observed prominent T-cell exclusion and increased TGFβ activity in the TME in primary tumours from orthotopically transplanted LAKTP MTOs (Extended Data Fig. 4c–f). TGFβ-activated stroma is a defining feature of poor prognosis consensus molecular subtype 4 (CMS4)[7]. Transcriptomic classification indicated that LAKTP MTOs displayed the CMS4 phenotype when transplanted in the caecum of syngeneic mice, but not when cultured *in vitro* (Fig. 1i and Extended Data Fig. 5). We obtained similar results with triple-mutant MTOs. These findings confirm that stromal gene expression is required to identify poor prognosis molecular subtypes[4,17–19] (see Supplementary Discussion).

Similar to microsatellite-stable (MSS) human CRCs[20], triple- and quadruple-mutant MTOs accumulated between 0.5 and 3.5 non-synonymous coding mutations per megabase (Extended Data Fig. 6a, b), which indicates that the genomes of these tumours evolved extensively. The most prevalent mutational signature in MTOs was signature 1, characteristic of MSS CRCs[21] (Extended Data Fig. 6c). MTOs and human MSS CRCs also displayed similar numbers of predicted high-affinity MHC-I-binding neoantigens, whereas mouse CRC cell lines CT26 and MC-38 exhibited around 40 times this number (Fig. 1j). Experimental liver colonization by MTOs caused an increased metastatic burden in the *nu/nu* background compared to wild-type C57BL/6J mice (Fig. 1k), indicating that the tumours were susceptible to T-cell-mediated adaptive immunosurveillance. Indeed,

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri i Reixac 10, 08028 Barcelona, Spain. [2]Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Barcelona, Spain. [3]Department of Pathology, Hospital del Mar, 08003 Barcelona, Spain. [4]Cancer Research Programme, Hospital del Mar Research Institute (IMIM), 08003 Barcelona, Spain. [5]Autonomous University of Barcelona (UAB), Spain. [6]ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain. [7]Department of Organic Chemistry, University of Barcelona, Martí i Franqués 1-11, 08028 Barcelona, Spain. †Present address: Cancer Research Programme, Hospital del Mar Research Institute (IMIM), 08003 Barcelona, Spain (A.C. and E.I.R.).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

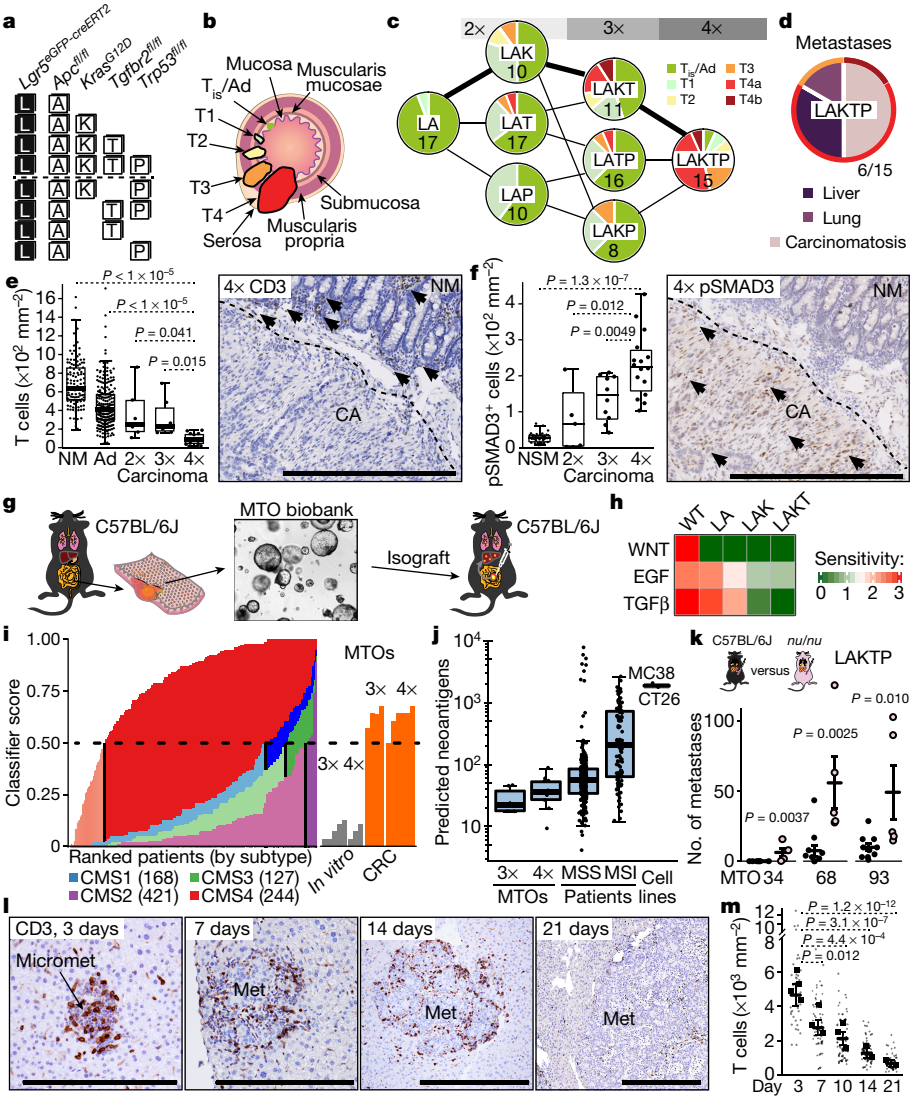


Figure 1 | Analysis of compound mutant mouse models and MTOs. **a**, Combination of alleles used to generate mouse models; dashed line distinguishes strains along the linear progression model[8] of A–K–T–P from other combinations of mutations. **b**, Classification of tumours according to the AJCC–TNM system, ranging from carcinoma *in situ* or adenoma (T_{is}/Ad) to four levels of increasingly invasive cancers (T1–T4, see Methods). **c**, Worst T-phenotype diagnosis per mouse, number of mice is indicated in the circles. The thick lines indicate the linear progression sequence. **d**, Metastasis in LAKTP mice. Outer ring, associated type of primary cancer. **e**, Box plot of T-cell density in normal mucosa (NM) and adenoma (Ad), compared to double-mutant (LAK, LAT and LAP, 2×), triple-mutant (LAKT, LAKP and LATP, 3×) and quadruple-mutant (LAKTP, 4×) cancers. $n = 119$, 242, 7, 8 and 16 regions of interest (ROIs), left to right. Right, CD3 immunohistochemistry of a LAKTP T4 carcinoma (CA), arrowheads indicate $CD3^+$ cells. Box plots have whiskers of maximum 1.5 times the interquartile range; the boxes represent first, second (median) and third quartiles. **f**, Box plot of $pSMAD3^+$ cell density in normal submucosal tissue (NSM), compared to 2×, 3× and 4× cancers, $n = 37$, 5, 10 and 16 ROIs, left to right. Right, pSMAD3 immunohistochemistry of a LAKTP T4 carcinoma, arrowheads indicate $pSMAD3^+$ cells. Two-sided Welch's *t*-test (**e**, **f**). **g**, MTO biobank and isograft technology. **h**, Mean sensitivity of MTOs to niche factors. WT, wild type. **i**, Left, cross-validation of consensus molecular subtype 4 (CMS4) classifier on human CRCs. Patients (shown as vertical bars) are ranked by subtype on the basis of classifier score. Right, classifier applied to 3× and 4× mutant MTOs *in vitro* (grey, $n = 5$ and 3, respectively) and to orthotopic isografts of those MTOs (orange, $n = 2$ and 3). **j**, Predicted MHC-I neoantigens in MTOs ($n = 6$ and 10, left and right, respectively), human CRCs ($n = 266$ and 112, left and right, respectively) and mouse CRC cell lines ($n = 2$). **k**, Liver metastasis generated by 4× MTOs in C57BL/6J or *nu/nu* mice ($n = 10$, 5, 11, 5, 10 and 5 mice, left to right; mean ± s.e.m.; two-sided Mann–Whitney *U*-test). **l**, CD3 immunohistochemistry on MTO129 liver metastases (Met) at the indicated days after intrasplenic injection. Micromet, micrometastasis. **m**, $CD3^+$ cell densities (individual metastases in grey and mean for mice in black; bars, group means ± s.e.m.); analysed with a mixed-effects linear model; $n = 4$, 3, 3, 2 and 2 mice, and $n = 39$, 58, 71, 56 and 56 tumours, left to right. Scale bars, 500 μm.

micrometastases were characterized by abundant $CD3^+$ T cells intermingled with tumour cells (Fig. 1l). Notably, T cells were progressively excluded at later time points (Fig. 1l, m).

Cell-population profiling of human or mouse CRC samples revealed that CAFs were the main contributors to TGFβ production (Extended Data Fig. 7a, b). We made use of the TGFBR1-specific inhibitor galunisertib[22] to inhibit TGFβ signalling in the TME. Treatment with galunisertib starting 11 days after transplantation of LAKTP MTOs in the caecum of syngeneic mice reduced primary tumour size, reduced the extent of carcinomatosis and blocked the appearance of liver metastases (Fig. 2a). Immunohistochemical quantification showed that galunisertib reduced the number of $pSMAD3^+$ cells (Fig. 2b) and gene expression profiling demonstrated decreased levels of TGFβ-response signatures[6] in fibroblasts (F-TBRS) and T cells (T-TBRS) (Fig. 2c).

To directly test therapeutic effects on liver metastatic disease, we inoculated LAKTP MTOs derived from either primary CRCs or liver

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

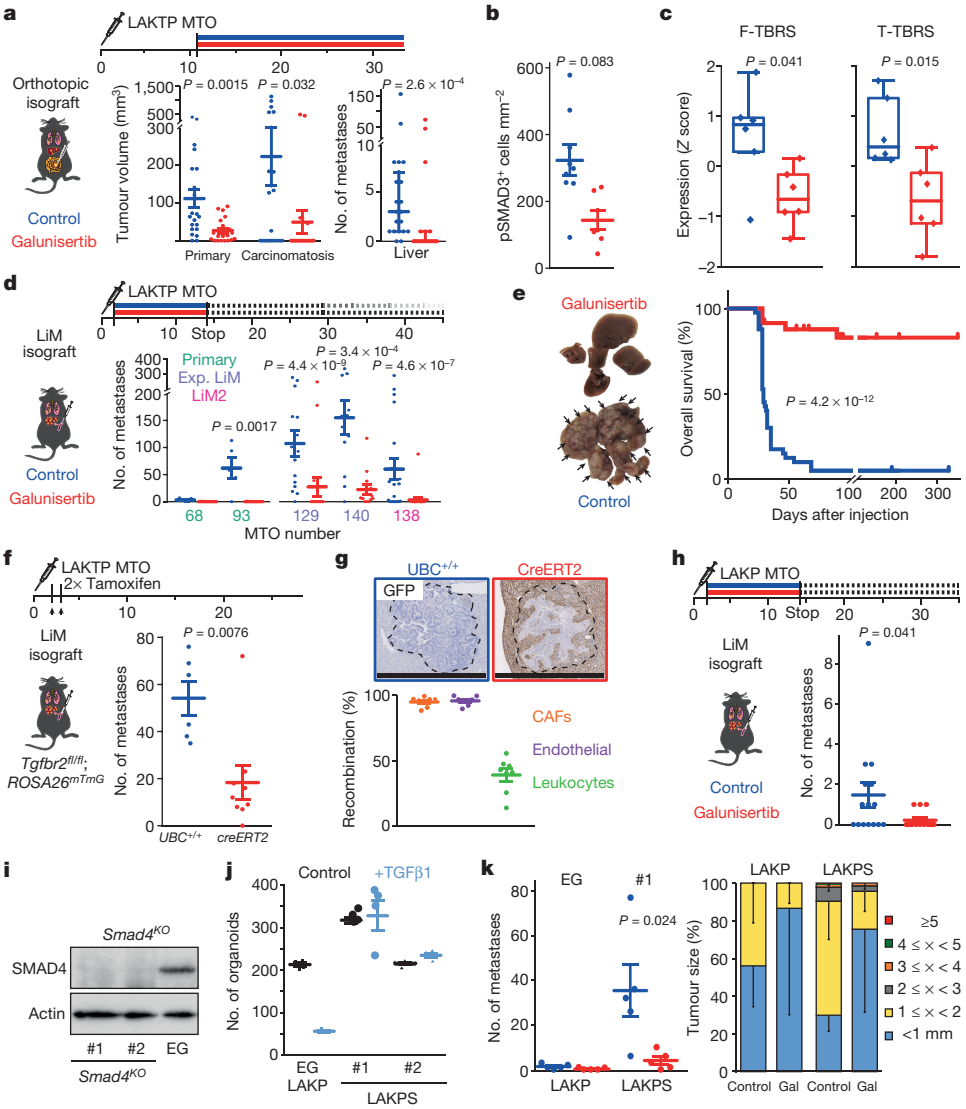


Figure 2 | Therapeutic effect of TGBβ inhibition. a, Galunisertib reduces tumour burden and metastasis in orthotopic MTO140 isografts. Middle, mean tumour volume ± s.e.m.; right, median number ± 95% confidence interval; $n = 21$ mice per condition, two-sided Mann–Whitney *U* test. **b**, pSMAD3$^+$ cell density (mean ± s.e.m.) in primary carcinomas ($n = 9, 7$); two-sided Student's *t*-test. **c**, Expression levels of TBRS in primary CRCs. Tukey box plots, $n = 6$ tumour samples, two-sided Mann–Whitney *U* test. **d**, Liver metastases (mean ± s.e.m.) generated by MTOs, treated with vehicle or galunisertib; $n = 5, 5, 5, 5, 15, 17, 12, 12, 20$ and 25 mice, left to right; analysed with a mixed-effects linear model. Exp. LiM, experimental liver metastasis; LiM2, second-round experimental liver metastasis. **e**, Representative livers at end point, arrows point to metastases. Kaplan–Meier survival curve of mice treated as in **d**. Control, $n = 41$ mice; galunisertib, $n = 35$ mice; Mantel–Cox test. **f**, Liver metastases (mean ± s.e.m.) in $UBC^{creERT2};Tgfbr2^{fl/fl}$ mice, recombined ($n = 9$) or not ($n = 6$). **g**, Liver tumours from **f**, stained for the recombination marker GFP, representative of three experiments. Scale bars, 1 mm. The graph shows the percentage of GFP$^+$ cells (mean ± s.e.m.). **h**, Liver metastases (mean ± s.e.m.) generated by 3× LAKP MTOs. $n = 15$ and 14 mice, left and right, respectively; two-sided Mann–Whitney *U* test. **i**, Western blot for SMAD4 and actin of LAKP empty guide (EG)- or CRISPR-targeted (#1 and #2) $Smad4^{KO}$ organoids. **j**, LAKP or LAKPS organoids. (mean ± s.e.m., $n = 4$ technical replicates) in control medium or treated with TGFβ1. **k**, Liver metastases (mean ± s.e.m.) from LAKP (EG) and LAKPS (guide #1) MTOs. Gal, galunisertib; *x*, diameter. $n = 5$ mice per condition; two-sided Mann–Whitney *U* test. Right, tumour diameters with s.e.m., $n = 10$ mice per condition.

metastases in the portal circulation; the latter exhibited enhanced metastatic capacity (Extended Data Fig. 6e). Treatment with galunisertib markedly decreased metastatic burden for all MTOs (Fig. 2d and Extended Data Fig. 6f), effectively curing a large proportion of mice (Fig. 2e). Galunisertib blocked TGFβ signalling in the TME of metastases (Extended Data Fig. 7c–h). We also transplanted LAKTP MTOs in mice bearing floxed *Tgfbr2* alleles that express a ubiquitous CreERT2 recombinase ($UBC^{creERT2};Tgfbr2^{fl/fl}$). Tamoxifen treatment induced recombination in fibroblasts, endothelial cells and leukocytes, and inhibited metastasis formation by isografted MTOs (Fig. 2f, g and Extended Data Fig. 8a). To analyse triple-mutant MTOs *in vivo*, we inoculated large numbers of cells. Four out of six tested MTOs (two LAKP and two LAKT) gave rise to metastases. However, these liver tumours remained very small (Extended Data Fig. 6g), possibly owing to niche factor dependencies[9,11,19]. Nevertheless, galunisertib treatment reduced liver metastases generated by LAKP MTOs (Fig. 2h and Extended Data Fig. 8b, c). We also introduced loss of function *Smad4* mutations (S) in LAKP MTOs using CRISPR–Cas9 technology. *In vitro*, LAKPS MTOs were resistant to the cytostatic action of TGFβ (Fig. 2i, j). *In vivo*, LAKPS formed larger numbers of liver tumours that were also larger in size than those formed by parental LAKP MTOs. Galunisertib effectively abrogated the metastatic capacity of LAKPS MTOs and, importantly,

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

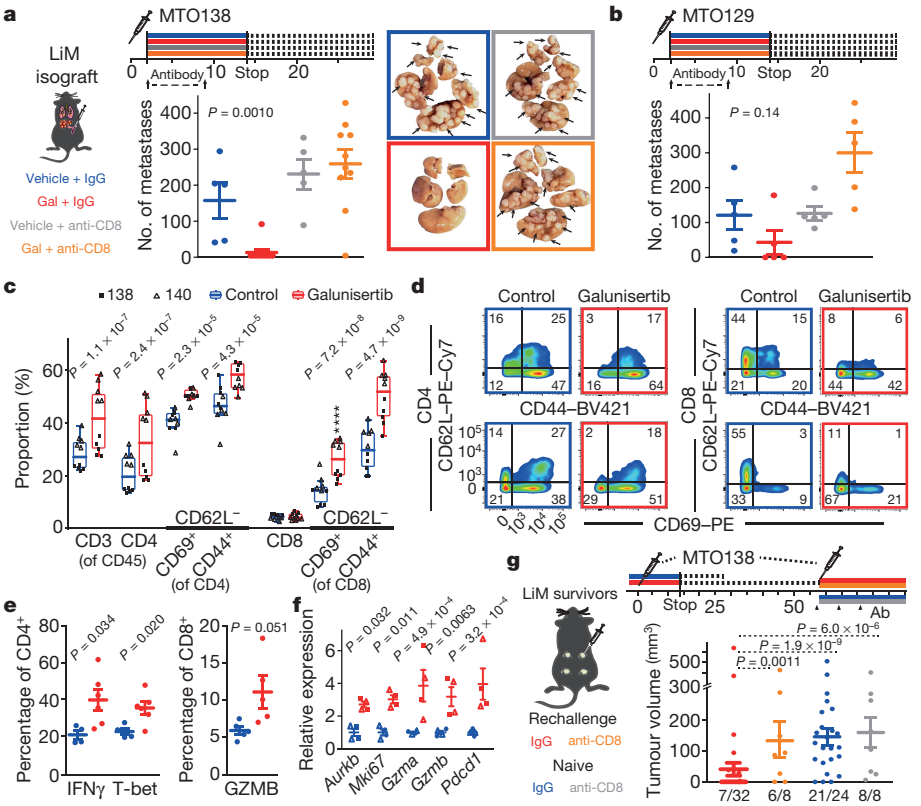


Figure 3 | TGFβ mediates immune evasion. a, **b**, Liver metastases (mean ± s.e.m.), $n = 5$, 10, 5, 10 mice, left to right (**a**), and $n = 5$ mice per condition (**b**). Two-sided Mann–Whitney *U* test. **c**–**e**, Immunophenotyping in whole livers from mice injected with MTO138 or MTO140 ($n = 5$ per condition). Tukey box plots, $n = 5$ mice per condition; analysed with a mixed-effects linear model. **d**, Flow cytometry plots from **c**, representative of two experiments. **e**, Intracellular cytokine expression (mean percentage ± s.e.m.), $n = 4$, 6, 4, 6, 5 and 5 mice, left to right; two-sided Student's *t*-tests. **f**, Relative mRNA expression (mean ± s.e.m.) on sorted CTLs from the experiment in **c**; $n = 4$ mice per condition; two-sided Student's *t*-test. **g**, Rechallenge experiment in liver metastasis survivors compared to naive hosts. Tumour volumes (mean ± s.e.m.) at end point. *x* axis, number of grown tumours/number of injections.

did not enhance growth of metastases generated by LAKP MTOs with wild-type TGFβ pathway (Fig. 2k and Extended Data Fig. 8d).

To study the kinetics of metastasis, we transduced MTOs with luciferase, which did not modify the therapeutic efficacy of galunisertib (Extended Data Fig. 8e). Bioluminescence revealed that galunisertib markedly enhanced cell killing at the onset of the exponential growth phase. However, this did not occur in *nu/nu* mice (Extended Data Fig. 8f–h). These kinetics are suggestive of an adaptive anti-tumour immune response. Indeed, the therapeutic effect of galunisertib was abolished upon depletion of $CD8^+$ cytotoxic T lymphocytes (CTLs) (Fig. 3a, b) or $CD4^+$ T-helper (T_H) cells (Extended Data Fig. 8i). We also transplanted MTOs in mice from the colony used to generate the original compound mutant mice. Galunisertib decreased metastatic burden in these hosts (Extended Data Fig. 8j), indicating that immunological rejection was not due to expression of exogenous antigens in MTOs such as those encoded by *eGFP* and *creERT2* transgenes.

We next assessed the status of the adaptive immune system in a transcriptomic dataset of human MSS CRC samples ($n = 981$) by measuring levels of expression signatures specific for naive or T_H1-activated $CD4^+$ T cells. Microsatellite-instable (MSI) CRCs displayed increased ratios of T_H1-to-naive T-cell genes, consistent with previous observations[23] (Extended Data Fig. 9a). In MSS CRCs, the ratio of T_H1 to naive cells was inversely correlated to the mean expression of *TGFB1*, *TGFB2* and *TGFB3* genes or the CAF-specific gene expression program (Extended Data Fig. 9b–f) and predicted disease relapse (Extended Data Fig. 9d, g). Therefore, abrogated T-cell differentiation, and increased TGFβ and CAF gene expression characterize a substantial subset of patients with MSS CRC and a poor prognosis.

Flow-cytometry analyses on whole livers colonized by MTOs for 7–10 days showed that there was increased recruitment of $CD3^+$ and $CD4^+$ cells after galunisertib treatment (Fig. 3c). This finding was confirmed by immunohistochemistry-based quantifications of liver sections, which also revealed an increased presence of T-bet, the master transcription factor regulating T_H1 cell differentiation (Extended Data Fig. 9h, i). Equivalent observations were made in treated liver metastases from parental MTOs, in experiments with orthotopic implantation, and during genetic abrogation of TGFβ signalling in the TME (Extended Data Fig. 10a–c). Importantly, both T_H cells and CTLs underwent activation, indicated by the increase in $CD44^+CD62L^-$ and $CD69^+CD62L^-$ populations (Fig. 3c, d), increased levels of T-bet and IFNγ effector molecules in T_H cells, and increased GZMB production and expression of proliferation genes in CTLs (Fig. 3e, f). We hypothesized that this robust immune response might immunize mice against the tumour cells. When survivors were rechallenged with the same MTO, most tumours were rejected within two weeks in the continued absence of treatment, whereas the MTOs grew efficiently in naive C57BL/6J recipients. Concurrent depletion of $CD8^+$ CTLs inhibited tumour rejection (Fig. 3g). We conclude that increased TGFβ levels in the TME limit adaptive immune responses by inhibition of the T_H1-effector cell phenotype.

Finally, we assessed the therapeutic action of galunisertib in mice with overt metastatic disease. Galunisertib treatment 14 days after MTO inoculation reduced metastatic burden but resulted in few complete remissions (Fig. 4a). Nevertheless, we observed increased infiltration of $CD4^+$ T cells and of T-bet^+ lymphocytes immediately after initiation of therapy (Extended Data Fig. 10d). Exploring resistance mechanisms to explain this mild response, we discovered that galunisertib-activated T cells exhibited marked surface expression of programmed cell-death protein 1 (PD-1) (Fig. 4b), and that metastases generated by MTOs progressively recruited stromal PD-1 ligand 1 $(PD\text{-}L1)^+$ cells—most

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

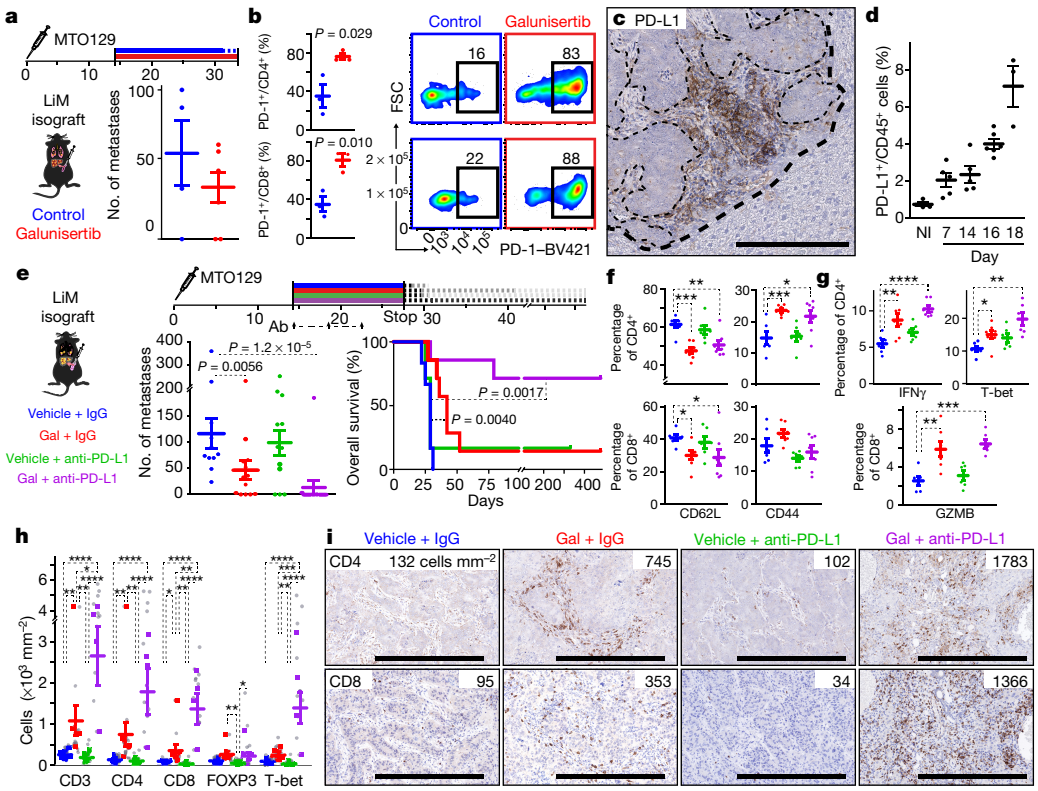


Figure 4 | Dual immunotherapy cures established metastases. **a**, Liver metastases (mean ± s.e.m.) in animals treated with galunisertib from day 14 after injection with MTO129, $n = 4$ and 6 mice, left and right, respectively. **b**, PD-1$^+$ T cells (mean ± s.e.m.) in microdissected liver metastases four days after the start of treatment (day 18), $n = 3$ per condition; two-sided Student's t-test. Right, representative density plots. **c**, PD-L1 immunohistochemistry on MTO129 liver metastases. **d**, PD-L1$^+$ leukocytes in liver metastases (mean ± s.e.m.) at indicated days after injection, or in non-injected (NI) liver, $n = 4, 5, 5, 6$ and 3, left to right. **e**, Treatment from day 14 of established liver metastases. Left, mean number of liver metastases ± s.e.m., $n = 11$, 13, 12 and 14 mice, left to right; two-sided Mann–Whitney U test; right, Kaplan–Meier survival curve; $n = 6$ (vehicle + IgG), 7 (galunisertib + IgG), 6 (vehicle + anti-PD-L1), 7 (galunisertib + anti-PD-L1), Mantel–Cox test. **f**, **g**, Surface activation markers (**f**) and intracellular markers (**g**) in mice euthanized two days after start of treatment (day 16). Mean percentage ± s.e.m., $n = 6, 7, 6, 7$ mice, left to right; two-sided Student's t-test. **h**, Cell densities quantified by immunohistochemistry from mice in **e**. Individual metastases in grey, mouse means in coloured squares; bars show group means ± s.e.m., $n = 5, 4, 4$ and 4 mice, left to right; analysed with a mixed-effects linear model. **i**, Images from the analysis in **h**; mean cell densities are indicated. Scale bars, 500 μm. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

prominently tumour-associated macrophages—as they expanded in size (Fig. 4c, d and Extended Data Fig. 10e, f). Treatment of mice bearing overt metastatic disease with anti-PD-L1 therapy alone had a very modest therapeutic effect (Fig. 4e). However, combined treatment with galunisertib and anti-PD-L1 antibodies induced a pronounced immune response, with increased T-bet and IFNγ levels in CD4$^+$ T_H cells and increased GZMB production in CTLs, which eradicated most metastases and prolonged recurrence-free survival for over a year after cessation of treatment (Fig. 4e–g). This response was associated with a marked, synergistic increase in infiltrating lymphocytes and T-bet expression (Fig. 4h, i), indicating both a disruption of the T-cell-exclusion phenotype characteristic of progressed metastatic disease and prominent T_H1 immune activation.

It has been hypothesized that MSS CRCs are immunologically 'cold', that is, scarcely T-cell infiltrated and possibly non-immunogenic, and that they are therefore unlikely to benefit from immune therapies[24] (Supplementary Discussion). By contrast, our data reveal that this class of CRCs can be killed effectively by the adaptive immune system through a CTL-dependent process, which CRC cells avert by increasing TGFβ levels. Consistent with the well-established role of TGFβ signalling in suppressing differentiation and activity of T cells[25–28], we observed that a TGFβ-activated TME antagonizes the T_H1-effector cell phenotype. We also show that such a TME excludes T cells from tumours, a phenomenon associated with poor outcomes across cancer types[3,29,30]. Enabling immune infiltration using TGFβ inhibitors is sufficient to confer susceptibility to anti-PD-1–PD-L1 checkpoint-based therapies, a strategy that may have broad applications for treatment of cancers that grow in a TGFβ-rich environment. These results strongly suggest that inhibition of TGFβ signalling could be promising as immunotherapy for patients with MSS, stroma-rich CRCs and a poor prognosis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 November 2016; accepted 2 January 2018.
Published online 14 February 2018.

1. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl Acad. Sci. USA* **105,** 4283–4288 (2008).
2. Mlecnik, B. *et al.* The tumor microenvironment and immunoscore are critical determinants of dissemination to distant metastasis. *Sci. Transl. Med.* **8,** 327ra26 (2016).
3. Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313,** 1960–1964 (2006).
4. Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* **47,** 320–329 (2015).
5. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).
6. Calon, A. *et al.* Dependency of colorectal cancer on a TGF-β-driven program in stromal cells for metastasis initiation. *Cancer Cell* **22,** 571–584 (2012).
7. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21,** 1350–1356 (2015).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

8. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61,** 759–767 (1990).
9. Fumagalli, A. *et al.* Genetic dissection of colorectal cancer progression by orthotopic transplantation of engineered cancer organoids. *Proc. Natl Acad. Sci. USA* **114,** E2357–E2364 (2017).
10. O'Rourke, K. P. *et al.* Transplantation of engineered organoids enables rapid generation of metastatic mouse models of colorectal cancer. *Nat. Biotechnol.* **35,** 577–582 (2017).
11. de Sousa e Melo, F.*et al.* A distinct role for Lgr5^{+} stem cells in primary and metastatic colon cancer. *Nature* **543,** 676–680 (2017).
12. Colnot, S. *et al.* Liver-targeted disruption of *Apc* in mice activates β-catenin signaling and leads to hepatocellular carcinomas. *Proc. Natl Acad. Sci. USA* **101,** 17216–17221 (2004).
13. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic *K-ras*. *Genes Dev.* **15,** 3243–3248 (2001).
14. Marino, S., Vooijs, M., van Der Gulden, H., Jonkers, J. & Berns, A. Induction of medulloblastomas in *p53*-null mutant mice by somatic inactivation of *Rb* in the external granular layer cells of the cerebellum. *Genes Dev.* **14,** 994–1004 (2000).
15. Levéen, P. *et al.* Induced disruption of the transforming growth factor beta type II receptor gene in mice causes a lethal inflammatory disorder that is transplantable. *Blood* **100,** 560–568 (2002).
16. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449,** 1003–1007 (2007).
17. Isella, C. *et al.* Stromal contribution to the colorectal cancer transcriptome. *Nat. Genet.* **47,** 312–319 (2015).
18. van de Wetering, M. *et al.* Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell* **161,** 933–945 (2015).
19. Fujii, M. *et al.* A colorectal tumor organoid library demonstrates progressive loss of niche factor requirements during tumorigenesis. *Cell Stem Cell* **18,** 827–838 (2016).
20. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
21. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).
22. Rodón, J. *et al.* First-in-human dose study of the novel transforming growth factor-β receptor I kinase inhibitor LY2157299 monohydrate in patients with advanced cancer and glioma. *Clin. Cancer Res.* **21,** 553–560 (2015).
23. Llosa, N. J. *et al.* The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. *Cancer Discov.* **5,** 43–51 (2015).
24. Le, D. T. *et al.* PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372,** 2509–2520 (2015).
25. Gorelik, L. & Flavell, R. A. Abrogation of TGFβ signaling in T cells leads to spontaneous T cell differentiation and autoimmune disease. *Immunity* **12,** 171–181 (2000).
26. Marie, J. C., Liggitt, D. & Rudensky, A. Y. Cellular mechanisms of fatal early-onset autoimmunity in mice with the T cell-specific targeting of transforming growth factor-β receptor. *Immunity* **25,** 441–454 (2006).
27. Thomas, D. A. & Massagué, J. TGF-β directly targets cytotoxic T cell functions during tumor evasion of immune surveillance. *Cancer Cell* **8,** 369–380 (2005).
28. Mehal, W. Z., Sheikh, S. Z., Gorelik, L. & Flavell, R. A. TGF-β signaling regulates CD8^{+} T cell responses to high- and low-affinity TCR interactions. *Int. Immunol.* **17,** 531–538 (2005).
29. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541,** 321–330 (2017).
30. Joyce, J. A. & Fearon, D. T. T cell exclusion, immune privilege, and the tumor microenvironment. *Science* **348,** 74–80 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Sancho for critical reading of this manuscript, all members of the Batlle laboratory for support and discussions, and members of the López-Bigas laboratory for help with revising the manuscript. We are grateful for the assistance of the IRB Barcelona core facilities for histopathology, functional genomics, mouse mutant and advanced digital microscopy; the flow cytometry and animal facilities of the UB/PCB; and the CRG genomic unit. D.V.F.T. held a Juan de la Cierva postdoctoral fellowship from MINECO. This work was supported by grants from the Doctor Josef Steiner Foundation, ERC advanced grant 340176, Instituto de Salud Carlos III, Olga Torres Foundation, BBVA Foundation, grant SAF-2014-53784 (MINECO) and by Fundación Botín. IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from MINECO.

Author Contributions D.V.F.T., S.P.-P., D.S. and A.Cal. performed animal husbandry and genotyping; M.S. performed immunohistochemistry; D.V.F.T. and M.I. analysed histopathology. D.V.F.T. generated MTOs, which D.V.F.T. and D.S. characterized *in vitro*; D.S. performed CRISPR experiments; C.S.-O.A. and A.B.-L. performed exome and RNA-seq analyses, other bioinformatics (CMS classifier and patient data), and statistical analyses. S.P.-P., D.V.F.T., J.B.-R., A.Cañ. and X.H.-M. performed mouse isografting; D.V.F.T., D.S. and J.B.-R. quantified immunohistochemistry. D.B., J.A.M. and A.R. synthesized galunisertib. D.V.F.T. coordinated and performed animal treatments and analysed the data. D.V.F.T., J.B.-R., S.I., E.I.R. and A.R.N. performed immunophenotyping experiments. E.B. and D.V.F.T. conceived the study, coordinated experiments and wrote the manuscript. E.B. supervised the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to E.B. (eduard.batlle@irbbarcelona.org).

Reviewer Information *Nature* thanks L. Vermeulen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

Ethics and animal maintenance. All experiments with mouse models were approved by the Animal Care and Use Committee of Barcelona Science Park (CEEA-PCB) and the Catalan government. Mice were maintained in a specific-pathogen-free (SPF) facility with a 12-h light–dark cycle and given *ad libitum* access to standard diet and water. All mice were closely monitored by the authors (D.V.F.T. and S.P.-P), facility technicians (during treatments) and by an independent veterinary scientist responsible for animal welfare.

Genetically modified mice. *Lgr5-EGFP-IRES-creERT2* (B6.129P2-*Lgr5*$^{tm1(cre/ERT2)Cle}$/J; stock 008875), *LSL-K-ras*^{G12D} (B6.129S4-*Kras*^{tm4Tyj}/J; stock 008179), *p53*^{LoxP} (B6.129P2-*Trp53*^{tm1Brn}/J; stock 008462) and *Tgfbr2*^{LoxP} (B6.129-*Tgfbr2*^{tm1Karl}/J; stock 012603) mouse strains[13–16] were obtained from The Jackson Laboratory, and *Apc*^{LoxP} mice were obtained from C. Perret[12]. Mice were inbred on a C57BL/6J background and successive crosses were performed to combine alleles. In this study, we have abbreviated the alleles: L (Lgr5$^{eGFP\ creERT2}$), A (*Apc*^{LoxP}), K (*LSL-Kras*^{G12D}), T (*Tgfbr2*^{LoxP}) and P (*p53*^{LoxP}) and generated the following strains: LAK, LATK, LAPK, LATPK. For practical reasons, the strains were maintained as homozygous (*lox/lox*) when possible (A, T, P) and we used both K-heterozygous and wild-type littermates. -*UBC*^{creERT2};*Tgfbr2*$^{fl/fl}$(B6.Cg-Tg(*UBC-creERT2*)1Ejb/2J; stock 008085) and Rosa26^{mTmG} (B6.129(Cg)-Gt(ROSA)26Sortm4 (*ACTB-tdTomato-eGFP*)Luo/J; stock 007676) mice have been described previously[31,32]. The latter allele allowed us to detect recombination using the shift from tdTomato to eGFP expression.

Recombination of genetic models. To induce tumorigenesis from mutated intestinal stem cells (ISC) with preferred localization in the distal part of the intestine, adult mice (at 12–13 weeks of age) were given drinking water (*ad libitum*) containing 2.5–3% (w/v) dextran sodium sulfate (DSS) for five days, and two intraperitoneal injections of diluted (4 mg/kg) tamoxifen (Sigma; dissolved in sunflower oil with 0.5% ethanol) on days 0 and 5. Thirteen per cent of mice treated in this way died in the first two weeks, most likely owing to DSS toxicity. These mice were excluded from the analysis. Animal weight, stool type and overall appearance were scored two times per week. With increasing morbidity, mice were evaluated more frequently until the experimental end point: progressive or rapid weight loss or emaciation and poor physical appearance, characterized by anaemia, hunched posture, ungroomed appearance and lethargy. Mice were then euthanized and dissected; the wall of the peritoneum was inspected and intestines, mesentery, liver, spleen, kidneys, diaphragm and lungs were collected. After macroscopic and/or binocular analysis, tissues were washed in phosphate-buffered saline (PBS), fixed in 10% formalin solution (Sigma) overnight and embedded in paraffin. Survival statistics were analysed using GraphPad Prism software (v.7.03). Median latency of metastasis was obtained by taking the median survival time of the animals in which metastases were detected.

Although the DSS-treatment-mediated induction of inflammation is acute and subsides after 2–3 weeks, we also induced gene recombination with tamoxifen without DSS treatment. Without this acute inflammation, induction still gave rise to tumours equivalent to those described here (including invasive carcinomas). However, tumour burden tended to concentrate to the distal ileum (small intestine), leading to adenomatous carpets, causing serious complications for the mouse, including reduced life span. The benefit of DSS is that it helps to target the colon. The experiments without DSS are not included in the survival and tumour assessment of this study.

Recombination on the *UBC*^{creERT2};*Tgfbr2*$^{fl/fl}$ background was performed with two shots of 80 mg/kg tamoxifen on days two and three. One mouse was excluded from the analysis shown in Fig. 2f because it exhibited no recombination upon tamoxifen treatment.

Histology and immunohistochemistry. Standard haematoxylin and eosin and antibody staining were performed on 4-μm tissue sections using standard procedures, as described previously[4]. Antibodies against CALD1 (rabbit; Sigma HPA008066; 1:250), IGFBP7 (rabbit; Sigma HPA002196; 1:200), pSMAD2 (rabbit; Cell Signaling 3108; 1:50), CD4 (rabbit; Sino Biological 50134-R001; 1:1000), CD8 (rabbit; Biorbyt orb10325-200; 1:200), FoxP3 (rabbit; Abcam ab54501; 1:1000), T-bet (Santa Cruz sc-21003; 1:500), pSMAD3 (rabbit; Abcam ab52903; 1:500) and PD-L1 (Cell Signalling 16764988S; 1:25) were used for staining overnight at 4 °C. Antibodies against GFP (rabbit; Life Technologies A11122; 1:500) and CD3 (rabbit; DAKO IS50330; 1:30) were used for staining for 2 h at room temperature. Images of histology and immunohistochemistry were taken with a Nikon Eclipse E600 and Nikon DS-Ri1 camera or with a Hamamatsu NanoZoomer Digital Slide Scanner (20× magnification).

Histopathological quantifications. Haematoxylin- and eosin-stained sections of intestines were scored blindly for T (tumour) status by an expert pathologist (M.I.) and a second observer (D.V.F.T.) using the AJCC–TNM system: scoring for T_{is} (carcinoma *in situ*), T1 (adenocarcinoma with submucosal invasion), T2 (intramuscular invasion), T3 (transmuscular invasion; serosa intact or cannot be evaluated) or T4 (transserosal invasion; T4b when local metastasis is evident, otherwise T4a). Across all genotypes, 156 tissue samples were scored (both small intestine and large intestine: caecum, colon and rectum) from 68 mice across eight genotypes. Given the high expression of *Lgr5* in the distal ileum and caecum, these tissues were in some cases overgrown with a carpet of dysplastic tissue, making adenoma counting impossible. However, we observed no obvious correlation between the percentage of dysplastic small-intestine surface and genotype. All invasive adenocarcinomas (small and large intestine) and large intestine adenomas were counted individually, reaching a total count of 1,477 tumours.

Mouse tumour organoid generation and culture. Before tissue fixation, tumours were dissected to include potential submucosal invasion and washed in PBS. Dissected material was roughly chopped with razor blades, followed by enzymatic digestion with 200 U/ml collagenase IV in DMEM (Life Technologies) for 20 min at 37 °C. Tissue fragments were subsequently treated by mechanical disruption in DMEM with 10% FBS (Life Technologies), washed with cold PBS and filtered through 100- and 40-μm meshes. Single-cell preparations were stained with propidium iodide (PI, a membrane integrity marker) and GFP^+PI^- cells were sorted in a FACS Aria flow cytometer (BD Biosciences). Typically, around 1,000 cells were obtained and plated in a drop of cold basement-membrane extract (BME) medium (Cultrex BME Type 2, Amsbio): 40 μl in a prewarmed (37 °C) well of a standard 24-well plate (Corning). After 5 min, mouse tumour organoid culture medium (advanced DMEM/F12, supplemented with 10 mM HEPES, Glutamax, B-27 without retinoid acid (all Life Technologies), 50 ng/ml recombinant human EGF (Peprotech), 100 ng/ml recombinant human NOGGIN and 1 μM galunisertib (LY2157299, see below)) was added. NOGGIN was produced in house; expressed as a His-tagged protein in HEK293-6E cells and purified with a 5-ml nickel-affinity column and a 5-ml HisTrap HP column in an ÄKTAxpress at 4 °C. Protein fractions were pooled and desalted (HiPrep 26/10 column), and tested on mouse *Apc*-mutant adenoma organoids or on bone morphogenic protein (BMP)-sensitive patient-derived organoids, analysing *ID1* and *ID3* gene expression by quantitative PCR. In initial passages, MTO medium was supplemented with the antimicrobial reagent Normocin (InvivoGen). MTOs were cultured at 37 °C with 85–90% humidity, atmospheric O_2 and 5% CO_2.

For MTO passaging, basement-membrane extract drops were washed once with HBSS (Lonza) and treated with Trypsin-EDTA (Sigma) for 20 min at 37 °C, followed by mechanical disaggregation of organoid fragments (by pipetting) until a single-cell suspension was obtained. Trypsin was quenched with FBS, cells were washed with HBSS and replated in cold BME medium on warm plates. MTOs were frozen as trypsinized organoids (single cells or small clusters) in DMEM with 50% FBS and 10% DMSO (Sigma). Cultures were checked bimonthly for mycoplasma contamination.

CRISPR genome editing. For gene knockout, guides were designed and cloned into px330-U6-Chimeric BB-CBh-hSpCas9 (Addgene: #42230), which was modified by introducing a SV40 promoter–IRFP expression cassette downstream of hSpCas9. Guide sequences (sgRNA indicated in upper case, restriction overhang in lower case):

Smad4: #1 forward, caccgAGACAGGCATCGTTACTTGT and reverse, aaacACAAGTAACGATGCCTGTCTc; *Smad4*: #2 forward, caccgAGTTTGATGTG TCATAGACA and reverse, aaacTGTCTATGACACATCAAACTc; *mCherryLuc*: forward, caccGCGCATGAACTCCTTGATGA and reverse aaacTCATCAAGG AGTTCATGCGC.

Organoids were nucleofected using a Nucleofector 2b (Lonza) in combination with the Cell Line Nucleofector Kit V (Lonza). Organoids were trypsinized for 15 min at 37 °C to generate single cells ($1.0–1.5 \times 10^6$ cells per guide), which were resuspended in 100 μl nucleofection buffer mix containing 3 μg sgRNA and nucleofected using program A32. Subsequently, cells were plated in BME medium and cultured in full growth medium. For *Smad4* knockout mutants, three days after nucleofection, growth medium was exchanged for selection medium (+ TGFβ1). Selection of *mCherry-Luc* knockout mutants was achieved by FACS sorting of *IRFP-mCherry* double-negative cells twice.

Western blotting. Trypsinized MTO cell pellets were resuspended in lysis buffer (1 mM EDTA, 1 mM EGTA, 1% SDS) containing protease inhibitor cocktail (Sigma). Protein concentration was determined using a standard Bradford assay (BioRad) and 30 μg per sample was separated by SDS–PAGE and transferred to a PVDF membrane (Millipore). Membranes were incubated with antibodies in TBS-Tween (0.2%) containing 5% milk at room temperature for 1 h. Antibodies against SMAD4 (B-8; Santa Cruz sc7966) and actin (Abcam ab20272) were used. The secondary antibody was horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG (Pierce 31430).

Immunohistochemical quantifications. Scanned CD3, CD4, CD8, FoxP3, T-bet and pSMAD3 immunohistochemistry stainings were analysed in QuPath (v.0.1.2) using the positive cell detection feature with empirical parameters. Several ROIs (tumours) were taken per section. In cases in which multiple sections per mouse or liver were considered, care was taken to avoid quantifying the same tumour

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

more than once. Data were processed and visualized with R and RStudio[33] (v.3.4.2 and v.1.1.383, respectively) and the ggplot2[34] package (v.2.2.1) (see Statistics and reproducibility).

Neoantigen prediction. Single-nucleotide variants that were annotated as non-synonymous or nonsense and in-frame insertions or deletions were selected for each sample. Peptides including ten bases upstream and downstream from the alteration were extracted from the canonical protein associated with the corresponding gene and were discarded if they matched another annotated protein. Human CRC: all mutations and predicted neoantigens for human CRC samples were downloaded from https://tcia.at/neoantigens. Samples were classified as MSS or MSI according to the TCGA consortium[5].

For both mouse and human data, neoantigens were predicted using netMHCpan software (v.2.8). See Supplementary Methods for a detailed description.

Mouse injections. For all injections, C57BL/6J (or athymic BALB/c *nu/nu*) mice were purchased from Janvier Labs at six weeks of age and injected at seven weeks. Sex was matched with the origin of the tumour. Intrasplenic or portal vein injections were used for liver colonization by the introduction of dissociated organoids (single cells) into the portal circulation; MTOs were cultured in standard six-well plates for four days and trypsinized as described in 'Mouse tumour organoid generation and culture'. The resulting cell suspension was filtered through 100- and 40-μm meshes (to remove clumps of cells and aggregated debris). Cells were counted and suspended in HBSS for injection, using $2–5 \times 10^5$ cells in 70 μl per mouse. Intrasplenic injections were performed as previously described[4,6,35]. For portal vein injections, a 30*G* syringe was used to inject 100 μl of cell suspension directly into the portal vein. Mice were euthanized at 3–5 weeks, visible liver metastases were counted and data were analysed using GraphPad Prism software (v.7.03). For subcutaneous injections, MTOs were cultured in standard six-well plates for three days and recovered with Cell Recovery Solution (BD Biosciences) for 40 min on ice. Organoids were washed in HBSS and a 10% fraction was trypsinized for cell counting (Neubauer or TC20 automated cell counter, Bio-Rad). Organoids were suspended in HBSS with 30% BME, an equivalent of 1.5×10^5 cells was injected in 2–4 locations on the flanks of each mouse. Growth was scored by palpation and volume was measured with a caliper (multiplying three orthogonal diameters and dividing by two) twice per week. In accordance with our IACUC protocol, mice with tumours over 300 mm^3 in volume were euthanized. Intracaecum injections were used for orthotopic isograft formation: MTOs were recovered as for subcutaneous injection, and an equivalent of 5×10^5 cells in 11 μl was injected per mouse. Intracaecum injections were performed using a 30*G* needle under binocular guidance, adapting the protocol used in ref. 36. Mice were euthanized ten weeks after injection and metastasis was scored macroscopically as well as histologically. Alternatively, orthotopic primary tumours were generated by transplantation of a 1-mm^3 piece of subcutaneously grown MTO-derived tumour (for example, for 3× MTO orthotopic transplantation); the piece was sutured on the tip of the caecum, which was folded over itself to mitigate carcinomatosis. Post-injection, all mice received analgesia (buprenorphine). The caliper technique was used to measure orthotopic tumour sizes after mice were killed. After measurement, pieces were dissected from fresh tumours and lysed in TRIzol (Life Technologies). RNA was purified with the Ambion PureLink kit (Life Technologies).

Mouse treatments. Galunisertib (also known as LY2157299) was synthesized in house and prepared as previously described[6]. Galunisertib or vehicle control was administered by gavage in a 0.15 ml volume, twice per day, starting two days after cell injection for metastasis-initiation experiments, unless otherwise indicated (treatments starting two days after injection give identical results to pre-treating two days before injection). A 800 mg/kg dose per treatment was used for all experiments except for the treatment of MTO138, or for the prolonged treatment of orthotopically implanted mice, in which 400 mg/kg doses were used. For *in vivo* $CD8^+$ T-cell depletion, rat CD8α (YTS 169.4; BioXCell BE0117) or rat IgG2b (LTF-2; BioXCell BE0090) isotype-control antibodies were diluted to 1 mg/ml in PBS and 200 μl was administered per mouse by intraperitoneal injection on days −1, 6 and 13 (unless otherwise indicated) relative to the day of MTO injection. For $CD4^-$ T_H cell depletion, rat anti-CD4α (GK1.5; BioXCell BE0003-1) or rat IgG2b isotype-control antibodies were used. For checkpoint immunotherapy or dual immunotherapy, we used rat PD-L1 (10F.9G2; BioXCell BE0101) or rat IgG2b (LTF-2; BioXCell BE0090) isotype-control antibodies.

***In vivo* study design.** Experimental group sizes were practically associated to cage sizes (five mice per cage) and treatment experiments were designed to have $n \geq 5$ mice per group (one or more cages). No mice were excluded from the analysis, unless explicitly indicated in the Methods. For gavage (galunisertib) treatment, as vehicle control and galunisertib are visually distinguishable, the only randomization we performed was the order of injecting mice: the researcher performing the injections was blinded to the treatment group. End-point criteria are equivalent to those described in 'Recombination of genetic models'. For genetic models, littermates were used as controls and cohorts were accumulated over time for logistical reasons. Mice were excluded if no tumours were observed after a year of survival post-induction.

Statistics and reproducibility. When no statistical control for confounders was needed, numbers of metastases in each condition were compared using a two-sided Mann–Whitney *U* test with exact computation of the null hypothesis, as implemented in the wilcox_test function in the R[33] coin package[37] or in GraphPad Prism. For Fig. 2d, where data were derived from different experimental runs and/or the same treatment performed on multiple MTOs, a mixed-effect linear model was fitted in which experiment run and/or MTO were included as a random effect.

In immunohistochemical experiments and the FACS experiment shown in Fig. 3c, group mean differences were assessed using a linear model. When the data involved more than one MTO (Fig. 3c and Extended Data Figs 9h and 10a) or more than one experiment run (Extended Data Figs 9h and 10c), these variables were included in the models as covariates. When multiple observations per mouse were derived, a mixed-effects linear model was fitted that included the mouse as a random effect (Figs 1m, 4h and Extended Data Figs 9h, 10a, c, d); otherwise, a standard linear regression model (only fixed effects) was used for analysis (Extended Data Fig. 10b).

In all cases, a fourth root was used to transform the outcomes in order to fulfil the assumptions of the linear models. Immunohistochemical quantifications: besides plotting individual tumours (grey), results were represented at the mouse level (coloured or black squares or MTO-dependent symbols) and at group level using the adjusted means and standard error of the means derived by the models in their original scale, after undoing the transformation. To achieve this, standard error of the means were computed by simulation from the corresponding model[38]. Wald tests derived from the linear models were used to perform pairwise comparisons between experimental conditions. A 5% level was chosen for statistical significance. All the analyses were performed using R[33]. Mixed-effects models were fitted using the R packages lme4[39] and lmerTest[40].

For the Kaplan–Meier survival curve after galunisertib treatment (Fig. 2e), all censored mice were metastasis-free when killed. Some mice were euthanized at the end point of control mice (severe morbidity), in order to count nodules at the same time point. Metastasis-burdened mice (≥1 tumour) were scored as death events, whereas disease-free mice were censored from the plot (ticks, zeroes in Source Data).

Tukey box plots in Figs 1e, f, j, 2c 3c and Extended Data Figs 3d, 7a and 9a–c, have whiskers of maximum 1.5 times the interquartile range; the boxes represent first, second (median) and third quartiles. In Fig. 1c, *P* values for comparison of phenotype severity, defined as severe (T3 or higher) versus less severe (<T3), are 0.0002 for LAKTP versus combined 3× genotypes, and <0.0001 for LAKTP versus combined 2× genotypes and for LAKTP versus LA (Fisher's exact test). We also tested for the presence (≥T1) or absence of carcinomas and the *P* values for these comparisons are: 0.0013 for LAKTP versus 3×, 0.0005 for LAKTP versus 2×, and <0.0001 for LAKTP versus LA (Fisher's exact test). The immunohistochemistry images in Fig. 1e, f are representative of four independent experiments. In Fig. 1i, numbers of patients are indicated; MTOs *in vitro*: 3×, $n = 5$ individual MTOs; and 4×, $n = 3$; plotted values are means of biological duplicates. MTOs orthotopically transplanted: 3×, $n = 2$ individual MTOs (MTO54 and 207); 4×, $n = 3$ individual MTOs (MTO68, 93 and 140); plotted values are biological duplicates ($2n$). The experiment in Fig. 1l, m was performed once. Besides providing individual *P* values, this model was used for the overall assessment of galunisertib treatment effect: control mean (95% confidence interval), 24.0 (10.5–47.7) versus galunisertib: 0.084 (0.00–0.81); $P = 1.2 \times 10^{-12}$. Data are from ten independent experiments. In Fig. 2e, data are from seven independent experiments, using either MTO129 or MTO138. In Fig. 2h, data are from MTO54 and MTO220, combined. Figure 2i, j, using MTO220, is representative of two independent experiments. Western blot source data are shown in Supplementary Fig. 1. In Fig. 2k, right, the values are fraction of total ± s.e.m. Data are from experiment with empty guide or from guides #1, #2 LAKP and S combined, from two independent experiments. The experiment in Fig. 3a was performed with MTO138, and the experiment in Fig. 3b was performed with MTO129. In Fig. 3g, *P* values refer to the comparison between: anti-CD8 and IgG in the survivors ($P = 0.0043$); rechallenge IgG and naive IgG, and between rechallenge IgG and naive anti-CD8 ($P < 0.0001$); two-sided Mann–Whitney *U* test. Figure 4b, c are representative of two experiments. In Fig. 4d, MTO129 is injected. In Fig. 4f, *P* values are for CD4 and CD62L: 0.0002 (control versus galunisertib) and 0.0051 (control versus double); CD44: 0.0008 (control versus galunisertib) and 0.0297 (control versus double); CD8 and CD62L: 0.0159 (control versus galunisertib) and 0.0181 (control versus double). In Fig. 4g, *P* values are for CD4 and IFNγ: 0.0099 (control versus galunisertib) and <0.0001 (control versus double); T-bet: 0.0208 (control versus galunisertib) and 0.0017 (control versus double); CD8 and GZMB: 0.0054 (control versus galunisertib) and 0.0001 (control versus double). For Fig. 4h, all *P* values are listed in Source Data. The experiment in Fig. 4i was performed once.

Patient data analysis. We used five publicly available Affymetrix microarray datasets from the NCBI GEO repository, which include gene expression and

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

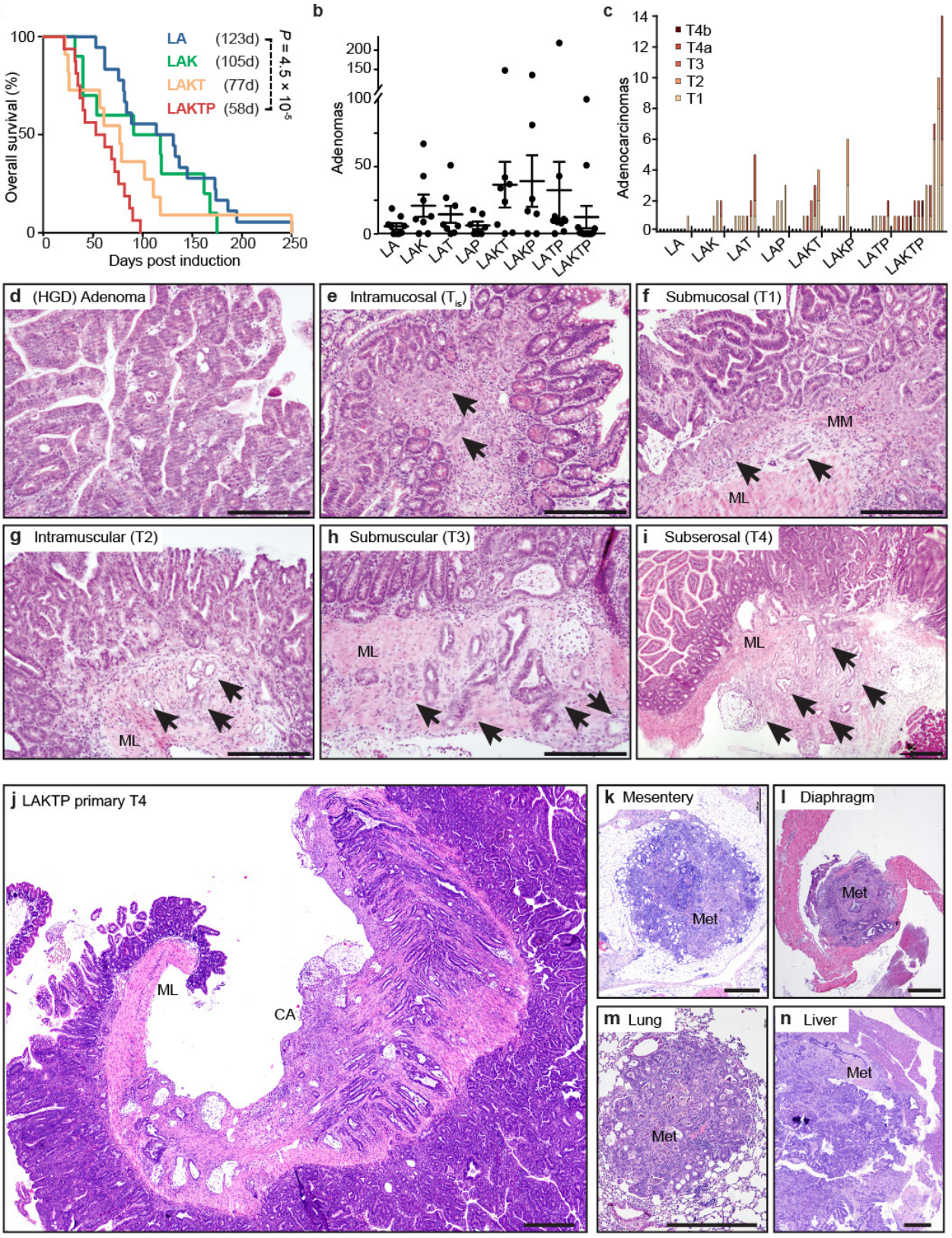
clinical information from a total of 1,194 patients with CRC. In order to obtain CD4-naive and $CD4^+$ T_H1 profiles, we carried out a differential expression analysis on gene expression dataset GSE22886[41]. We defined the CD4 naive signature to include genes that are upregulated at least threefold in both CD4-naive versus T_H1 and in CD4-naive versus Th2; Benjamini–Hochberg false discovery rate[42] (FDR) < 5% in both comparisons. Similarly, the CD4-T_H1 signature included genes that were upregulated at least threefold compared to CD4-naive, and limma FDR < 5%. Association with metastasis was evaluated using a frailty Cox proportional hazards model. Statistical significance was assessed by means of a log-likelihood ratio test, and Wald tests were used for pairwise comparisons when necessary. Association of expression intensities was evaluated as continuous variables assuming a linear relationship with the logarithm of the relative risk. Sample groups of low, medium and high expression levels were defined using the tertiles of the intensity distribution after correction for technical effects. For visualization purposes, Kaplan–Meier curves were estimated for groups of tumours showing low, medium and high gene or signature expression. Only samples from patients diagnosed in stages I, II and III were taken into consideration for these analyses. The threshold for statistical significance was set at 5%. All analyses were carried out using R[33,43]. See Supplementary Methods for a detailed description.

Classification of mouse tumours according to consensus molecular subtypes. We used the *k*-top scoring pairs (*k*-TSP) algorithm and trained the classifier for a binary outcome (CMS4 versus not-CMS4) in 960 samples. See Supplementary Methods and Extended Data Fig. 5 for a detailed description.

Data availability. MTO whole-exome sequencing and RNA sequencing analysis (RNA-seq) data have been deposited in the ENA archive with accession number PRJEB22559. Expression array data are available at GEO with accession number GSE103562. Computer code is available upon request. Source Data are available in the online version of the paper.

31. Ruzankina, Y. *et al.* Deletion of the developmentally essential gene *Atr* in adult mice leads to age-related phenotypes and stem cell loss. *Cell Stem Cell* **1,** 113–126 (2007).
32. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45,** 593–605 (2007).
33. R Core Team. *R: a Language and Environment for Statistical Computing* http://www.r-project.org (R Foundation for Statistical Computing, 2017).
34. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
35. Warren, R. S., Yuan, H., Matli, M. R., Gillett, N. A. & Ferrara, N. Regulation by vascular endothelial growth factor of human colon cancer tumorigenesis in a mouse model of experimental liver metastasis. *J. Clin. Invest.* **95,** 1789–1797 (1995).
36. Céspedes, M. V. *et al.* Orthotopic microinjection of human colon cancer cells in nude mice induces tumor foci in all clinically relevant metastatic sites. *Am. J. Pathol.* **170,** 1077–1085 (2007).
37. Hothorn, T., Hornik, K., Van der Wiel, M. A. & Zeileis, A. A Lego system for conditional inference. *Am. Stat.* **60,** 257–263 (2006).
38. Gelman, A. & Hill, J. Data analysis using regression and multilevel/hierarchical models. *J. Stat. Softw.* **30,** 1–5 (2009).
39. Bates, D., Maechler, M., Bolker, B. M. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67,** 1–48 (2015).
40. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: tests in linear mixed effects models *J. Stat. Softw.* **82,** 1–26 (2015).
41. Abbas, A. R. *et al.* Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun.* **6,** 319–331 (2005).
42. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300(1995).
43. Gentleman, R., Carey, V., Huber, W., Irizarry, R. & Dudoit, S. (eds) *Bioinformatics and Computational Biology Solutions using R and Bioconductor.* https://doi.org/10.1007/0-387-29362-0 (Springer, 2005).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

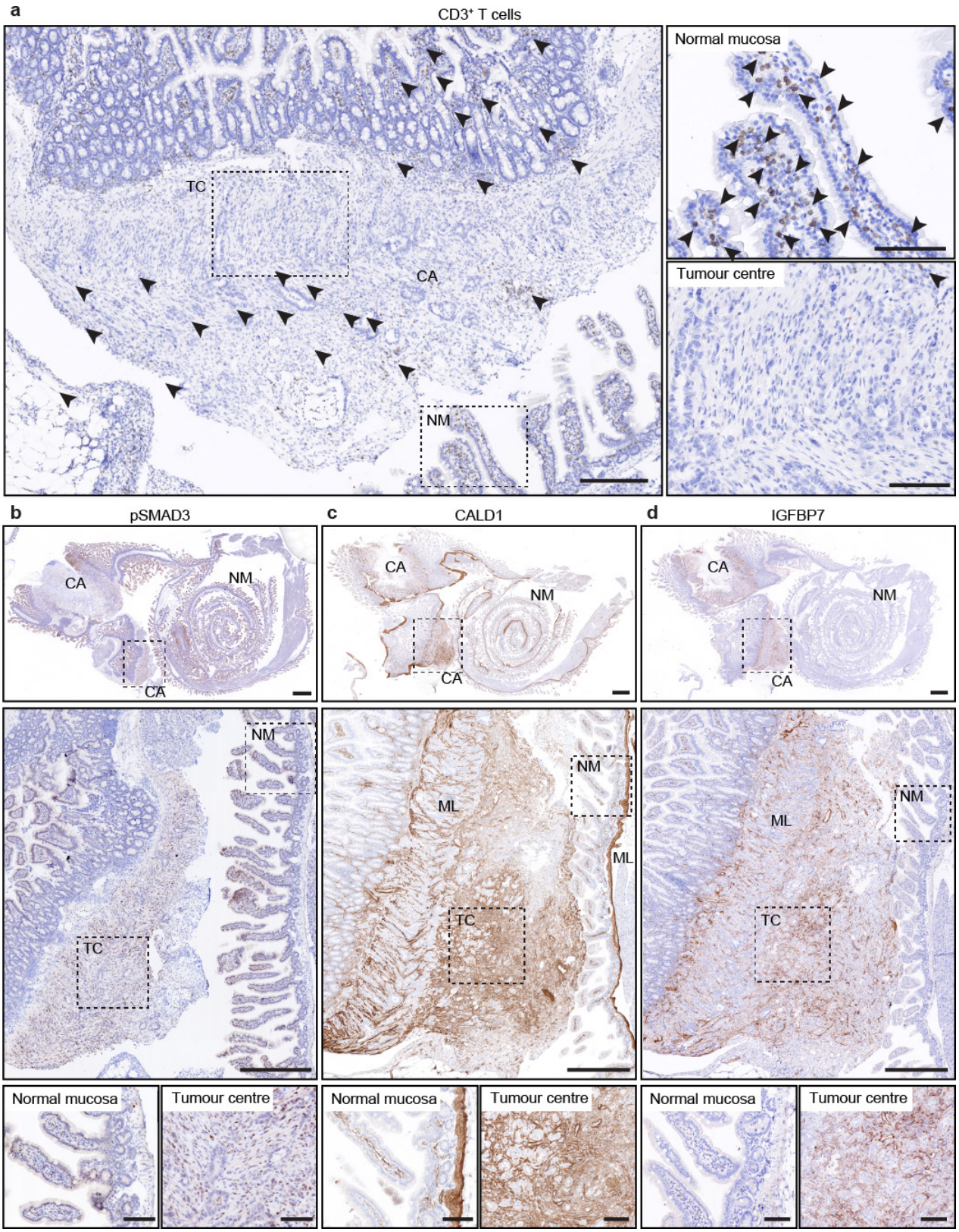


Extended Data Figure 1 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

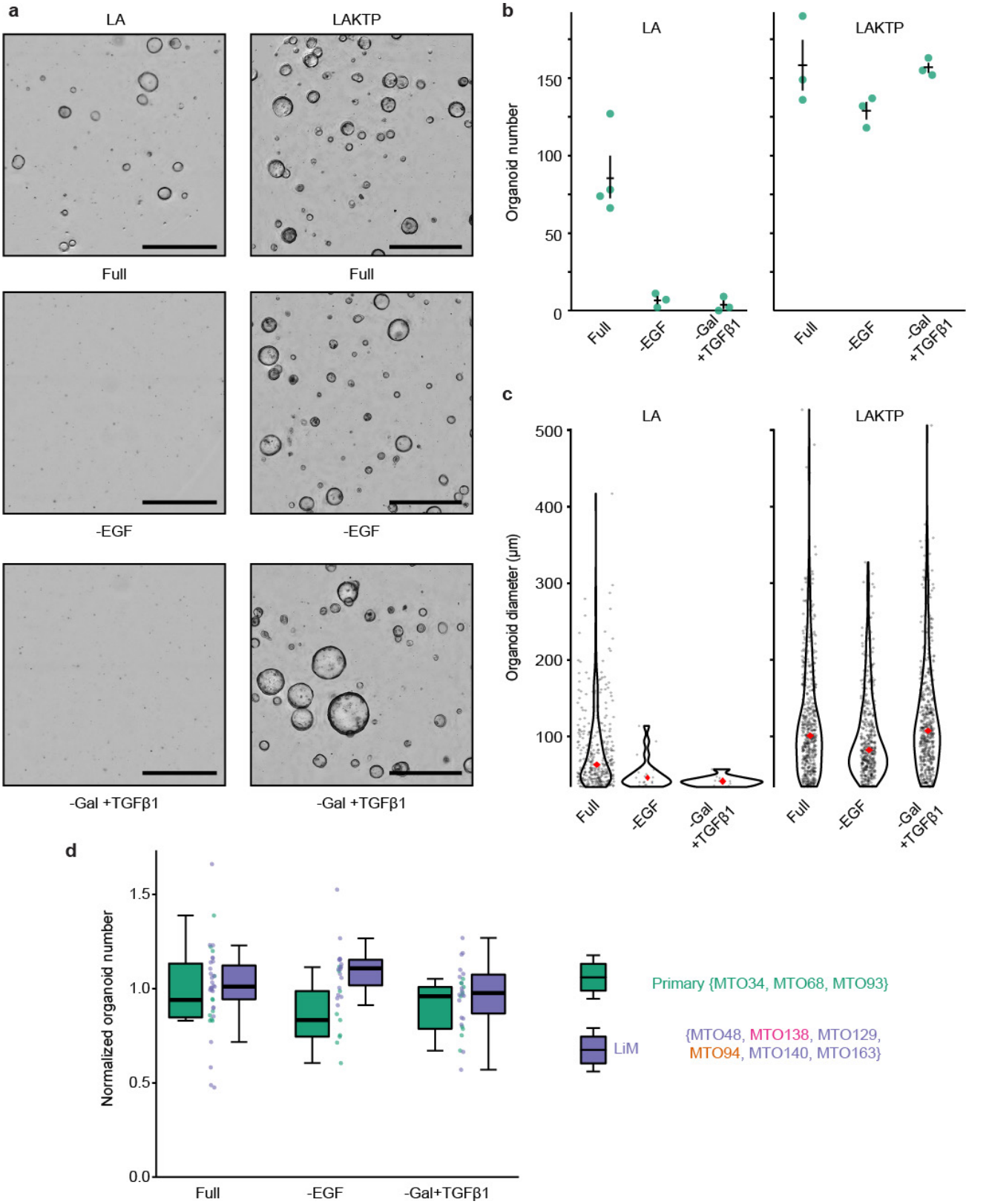
Extended Data Figure 1 | Additional description of the genetic mouse models. a, Kaplan–Meier survival curves for LA ($n = 18$ mice), LAK ($n = 10$), LAKT ($n = 11$) and LAKTP ($n = 16$) mice, as days post-tamoxifen induction of recombination in intestinal stem cells. Median survival is indicated in parentheses; Mantel–Cox test. **b**, Number of tumours (adenomas and carcinomas) in the large intestine, by genotype. Numbers for individual mice ($n = 9$ (LA), 8 (LAK, LAT and LAKT), 7 (LAP and LAKP), 10 (LATP), and 13 (LAKTP)) are indicated, as well as mean ± s.e.m. **c**, Number and grade of adenocarcinomas per individual mouse (bars), by genotype. *P* value for pairwise comparisons of tumour number: LA versus 2×, 0.0198; LA versus 3×, 0.0337; LA versus LAKTP, 0.0001; 2× versus LAKTP, 0.0051; and 3× versus LAKTP, 0.0014; $n = 9$ mice (LA), 8 (LAK, LAT, LAKT), 7 (LAP, LAKP), 10 (LATP), 13 (LAKTP); two-sided Mann–Whitney *U* test. **d**–**i**, Example images of tumour types scored by the TNM criteria, representative of 23 independent experiments. Arrows indicate invasive cells or glands, absent in an adenoma (**d**), restricted to the mucosa in an *in situ* tumour (Tis, **e**), breaching the muscularis mucosae (MM) in a submucosal invasive tumour (T1, **f**), penetrating the muscle layer (ML) in an intramuscular invasive tumour (T2, **g**), reaching the serosa layer beyond the ML in a submuscular invasive tumour (T3, **h**), or infringing all layers including the serosa in a subserosal or T4 tumour (**i**). **j**, Representative micrograph of an haematoxylin- and eosin-stained invasive subserosal adenocarcinoma (T4) of an LAKTP mouse. Carcinoma (CA) glands are indicated, invading through the muscle layer (ML). **k**–**n**, Examples of spontaneous metastases (Met) in the mesentery, diaphragm, liver and lung of induced LAKTP mice, each image is representative of two experiments. Scale bars, 500 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



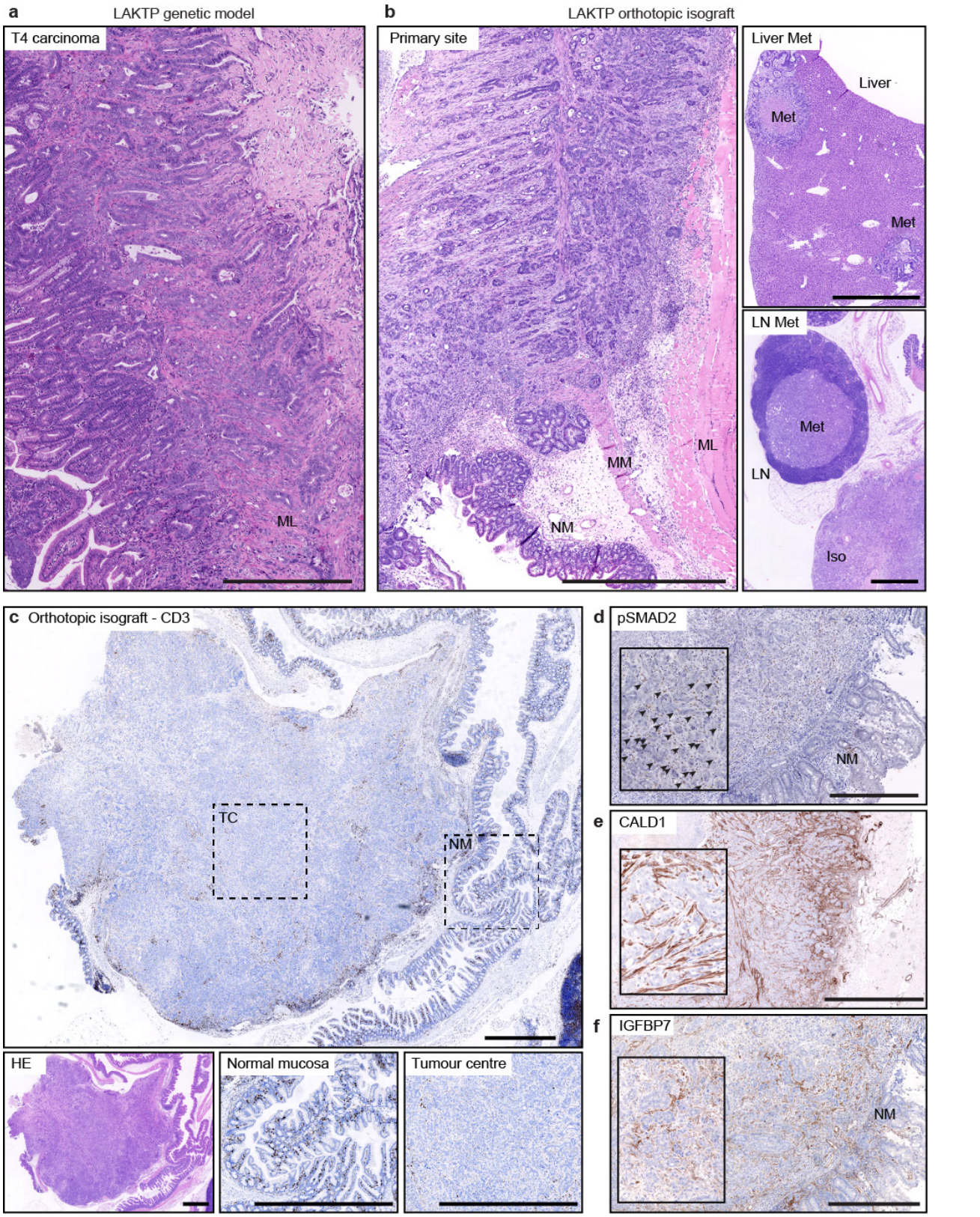
Extended Data Figure 2 | Analysis of the TME of LAKTP adenocarcinomas in the genetic model. a, Immune infiltration in LAKTP carcinomas. Immunohistochemistry for CD3, indicating infiltrating T cells, in an intestinal adenocarcinoma of an LAKTP mouse. Insets show a magnified view of normal mucosa and tumour centre (TC). Arrowheads indicate T cells. **b**–**d**, Markers of TGFβ activation and poor prognosis in LAKTP carcinomas in a representative section of intestine with two aggressive invasive adenocarcinomas. **b**, Staining for pSMAD3. Note nuclear staining in stromal cells of the tumour centre. **c**, Staining for CALD1. Note that it is expressed in the muscle layer throughout the intestine, and in a subset of cells in the stroma of the normal mucosa. In the tumour centre of invasive carcinomas, the staining intensity is clearly higher, especially in fibroblast-like cells. **d**, Staining for IGFBP7, which is barely detectable in the normal mucosa. Expression is strong in the TME of the tumour centre. Images are representative of four independent experiments Scale bars, 1 mm (**b**–**d**, top); 500 μm (**a**, left; **b**–**d,** middle); 100 μm (all other panels).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



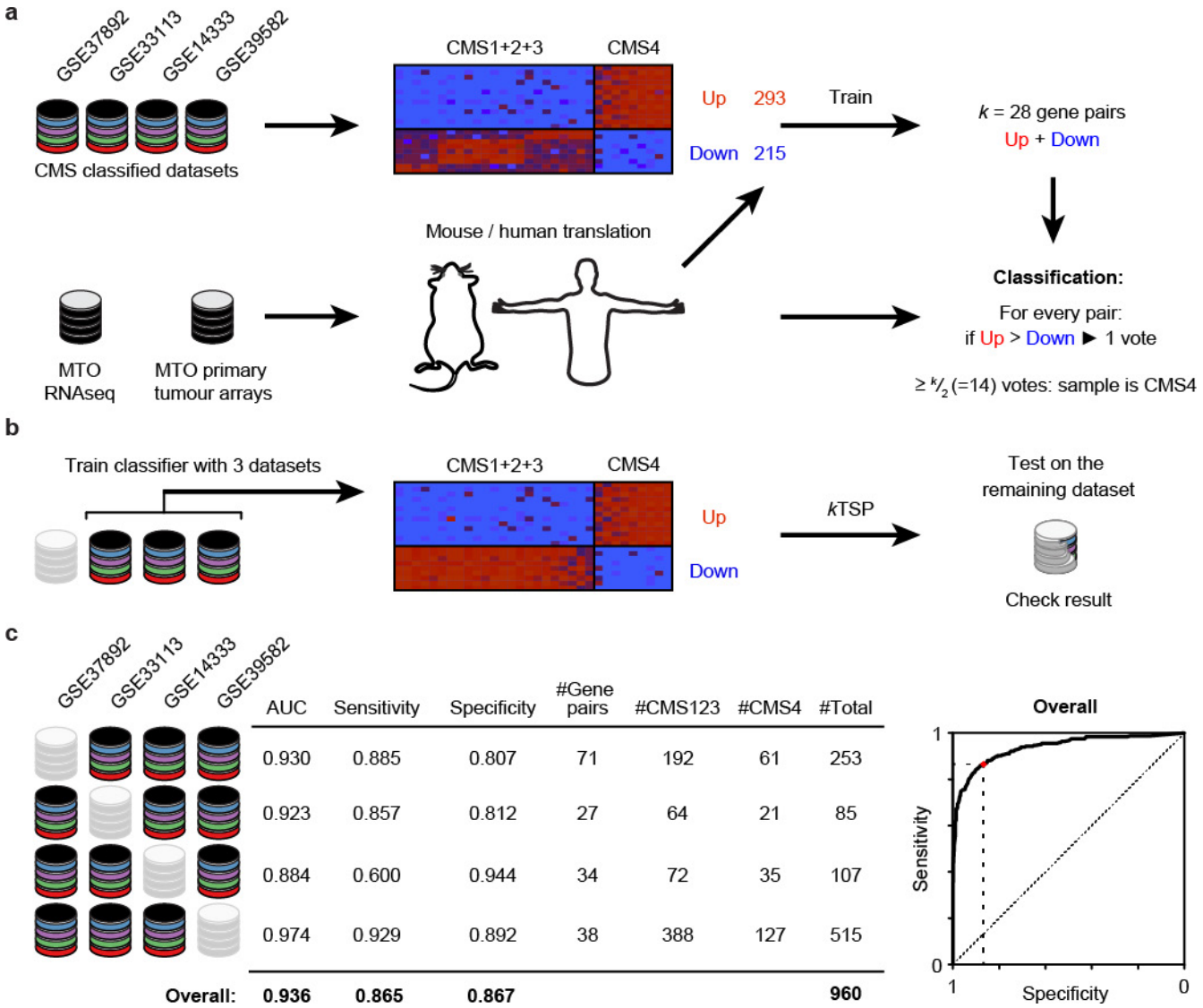
Extended Data Figure 3 | Niche factor analysis of MTOs *in vitro*. Organoid formation assay in different media with representative MTOs from LA and LAKTP mice. **a**, Images obtained by automated plate scanning, *z*-stack projection and stitching of the niche factor requirement assay. Shown are full medium, medium without EGF and medium without galunisertib (Gal) but with TGFβ1. **b**, Results from automated organoid detection for the MTOs shown in **a**, each data point shows the number of organoids in three replicate wells for representative analyses. Data are shown as mean ± s.e.m. **c**, Organoid diameter was calculated from detected pixel area. Data points are individual organoids, pooled from triplicate wells. Violin plots and medians are shown. **d**, Summary of all data obtained from different LAKTP MTOs analysed in triplicates, from both primary tumours and liver metastases. Data (individual values and Tukey box plots) are normalized to the mean of full medium. Scale bars, 1 mm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 4 | Histology and TME of primary tumours from orthotopically isografted LAKTP MTOs. a, **b**, Haematoxylin and eosin staining of a LAKTP-T4 carcinoma in the genetic model (**a**) and of a recapitulated adenocarcinoma developed from orthotopically isografted LAKTP MTO93 (**b**). Liver and lymph node (LN) metastasis observed in orthotopic isografted mouse (**b**, right panels). Iso, primary isograft. **c**–**f**, Markers of poor prognosis in isografted primary tumours. **c**, Immunohistochemistry for the T-cell marker CD3 (**c**, main panel and right, bottom right two panels) and haematoxylin and eosin stain (**c**, bottom left) of an orthotopic isograft, with typical glandular differentiation. Note that there are few infiltrating T cells in the tumour centre (**c**, bottom right), compared to the periphery and the normal mucosa (**c**, middle right). **d**–**f**, An orthotopically isografted, invasive primary tumour, stained for pSMAD2 (**d**, arrows indicate positive stromal cells), CALD1 (**e**) and IGFBP7 (**f**). Images represent three independent experiments. Scale bars, 1 mm (**c**, main and bottom left); 500 μm (all other panels).

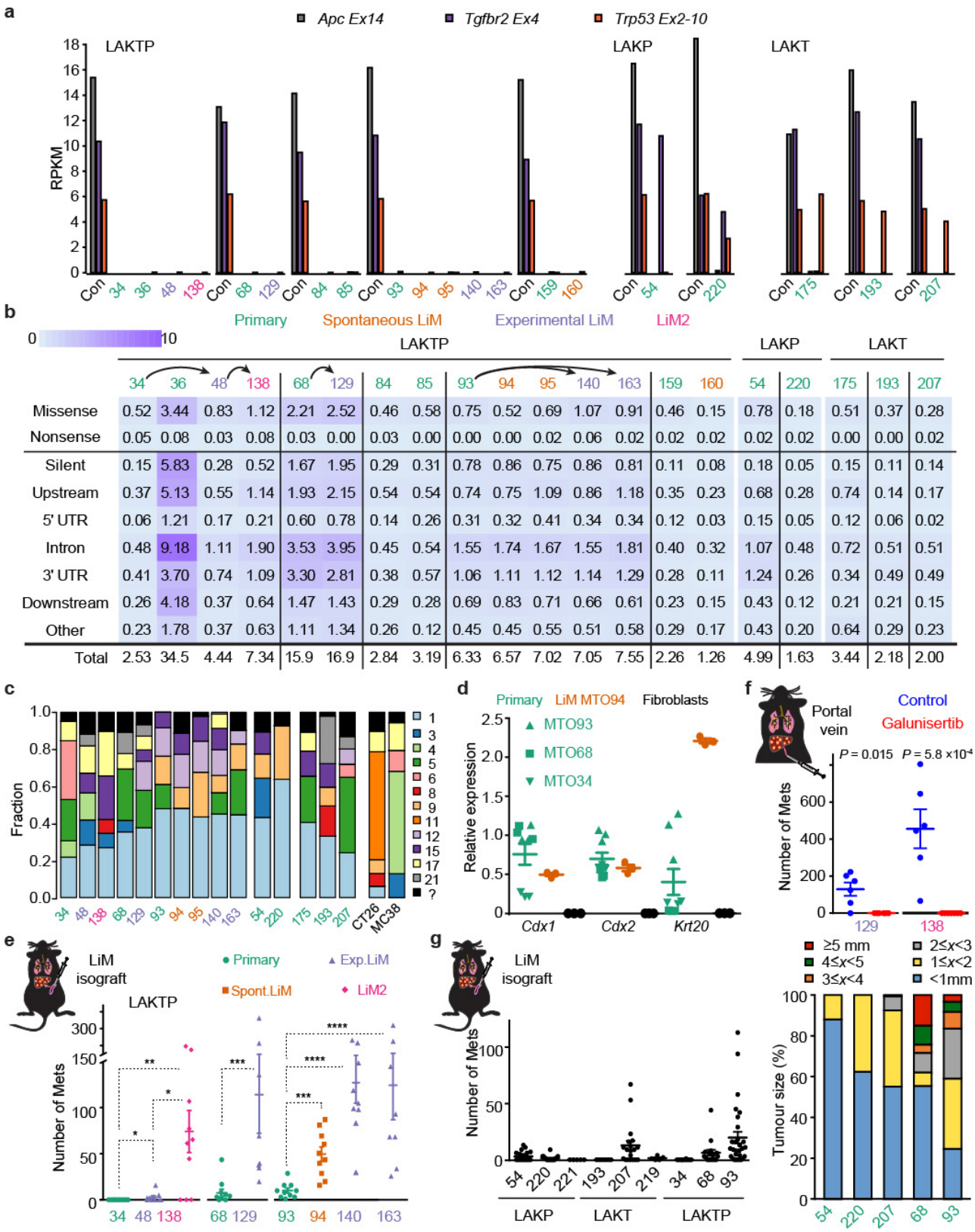
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



AUC	Sensitivity	Specificity	#Gene pairs	#CMS123	#CMS4	#Total
0.930	0.885	0.807	71	192	61	253
0.923	0.857	0.812	27	64	21	85
0.884	0.600	0.944	34	72	35	107
0.974	0.929	0.892	38	388	127	515
Overall:	**0.936**	**0.865**	**0.867**			**960**

Extended Data Figure 5 | Schematic of the CSM4 classifier, using the *k*-TSP algorithm. a, A metacohort of four pooled human CRC datasets with CMS annotation is used to find upregulated and downregulated genes in the CMS4 subtype (434 with >1.5-fold change and 300 with <0.8-fold change). To ensure usability in mouse samples, candidate genes are filtered by high-confidence homology between the two species. The *k*-TSP algorithm selects from among the up- and down-regulated genes the optimal set of pairs that correctly classifies CMS4 samples. The resulting classifier is applied to mouse samples by comparing the expression of pairs of genes. For each sample, the score is increased by one unit if the expression of the 'Up' gene is greater than that of the 'Down' gene. The maximum score for CMS4 classification is 14. **b**, Cross-validation by a leave-one-out approach. To assess performance, we repeatedly trained a classifier on three out of the four human datasets and tested using the remaining dataset. **c**, Results of the leave-one-out analysis. Area under the curve (AUC) was computed after re-scaling the classifier scores to the range between zero and one. Right, receiver operating characteristic curve.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

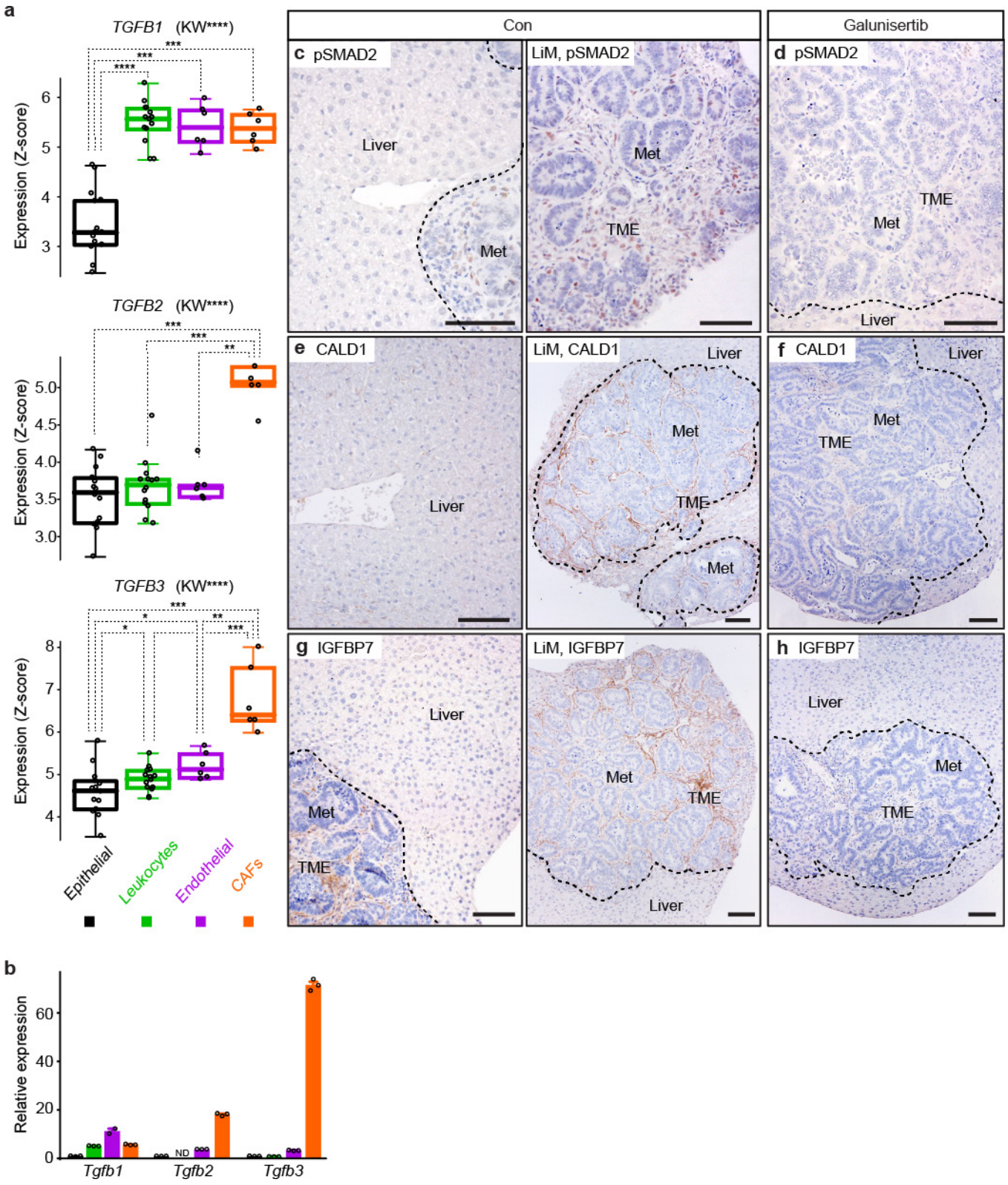


Extended Data Figure 6 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 6 | Genetic and functional analyses of MTOs. **a**, Whole-exome sequencing of MTOs. Number of reads for the recombined exons in reads per kilobase per million mapped reads (RPKM) in MTOs and their matched host tail DNA (Con). Mouse genotypes are indicated. **b**, Overview of the different LAKTP MTOs analysed, including experimental metastatic descendance (indicated by arrows), and heat map and table of the number of indicated type of somatic mutation per Mb of sequenced exome. **c**, Mutational signatures[21] in the MTOs and mouse monolayer cell lines CT26 and MC38. As in human MSS CRC, mutations in MTOs (but not in the cell lines) are predominantly of type 1. **d**, mRNA expression of intestinal epithelial genes *Cdx1*, *Cdx2* and *Krt20* in cultured MTOs derived from primary tumours MTO34, 68 and 93, and spontaneous liver metastasis MTO94, compared to mouse liver fibroblasts. Values are relative to $2^{-\Delta Ct}$ for *Ppia*, normalized to the value for MTO93; and are shown as three technical replicates with mean ± s.e.m. **e**, Number of liver nodules after intrasplenic injection of primary LAKTP tumour and spontaneous or experimental liver metastasis (LiM)-derived MTOs in syngeneic C57BL/6J mice. Data points represent individual mice; data are mean ± s.e.m. *P* values (95% confidence interval) are MTO34 ($n = 10$ mice) versus MTO48 ($n = 9$): 0.0325 (−2–0), MTO48 versus MTO138 ($n = 10$): 0.0305 (−108–0); MTO34 versus MTO138: 0.0031 (−108–0); MTO68 ($n = 11$) versus MTO129 ($n = 8$): 3.2×10^{-4} (−207 to −30); MTO93 ($n = 10$) versus MTO94 ($n = 10$): 1.1×10^{-4} (−60 to −19); MTO93 versus MTO140 ($n = 9$): 2.2×10^{-5} (−149 to −81); MTO93 versus MTO163 ($n = 8$): 9.1×10^{-5} (−220 to −31); two-sided Mann–Whitney *U* test. **f**, Number of liver nodules (mean ± s.e.m.) after inoculation of disaggregated MTO129 or MTO138 directly into the portal vein in syngeneic animals. Data are from individual mice, $n = 6$ per condition. 95% confidence interval: MTO129, (−223–0) and MTO138 (−805 to −66); two-sided Mann–Whitney *U* test. **g**, Number of liver nodules (mean ± s.e.m.) of intrasplenically injected 3× and 4× MTOs, $n = 22$ mice (MTO54), 20 (MTO220), 5 (MTO221), 7 (MTO193), 19 (MTO207), 5 (MTO219), 14 (MTO34), 16 (MTO68) and 29 (MTO93). Right, distribution of tumour diameters.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

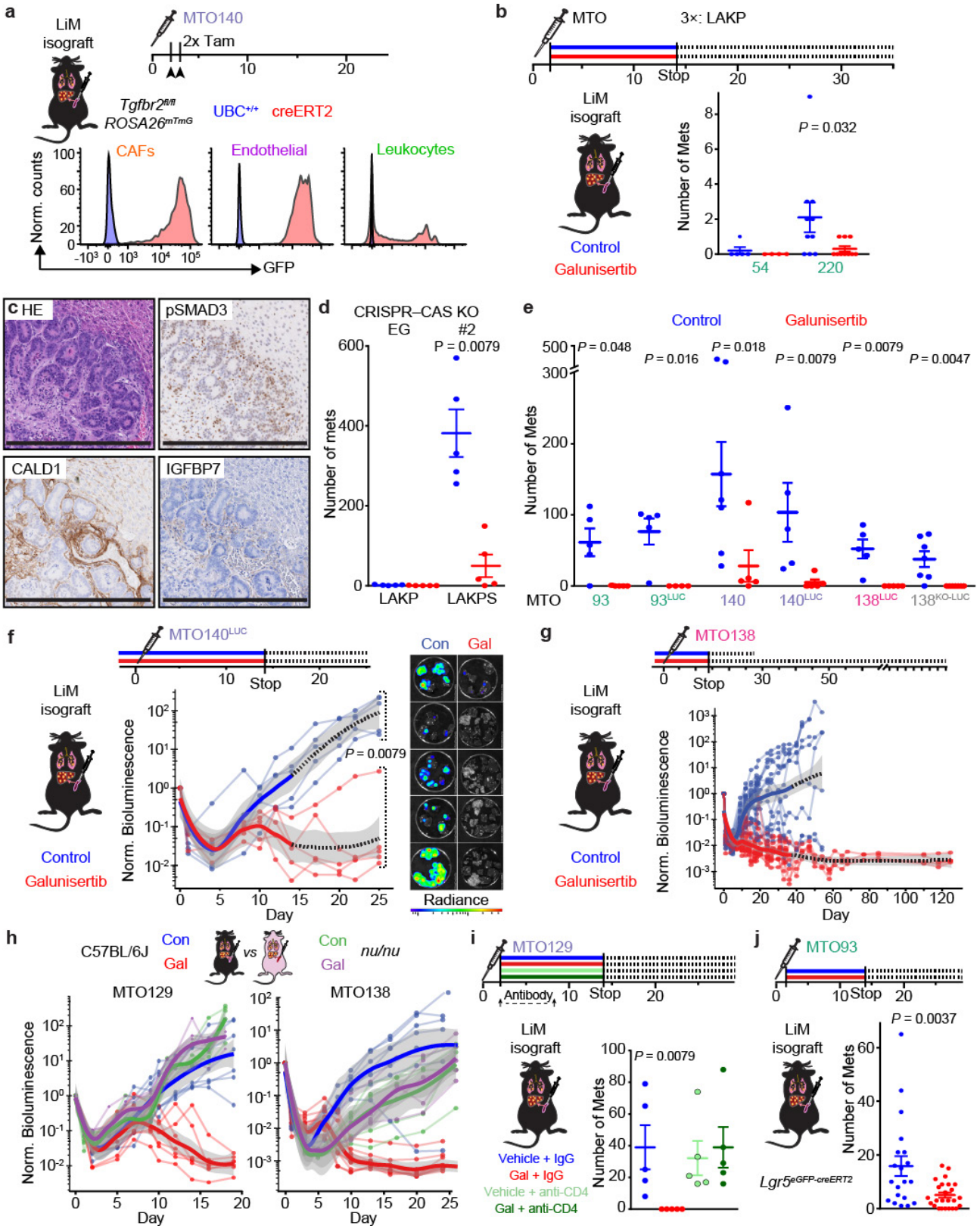


Extended Data Figure 7 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 7 | Source of tumoral TGFβ and pathway blockade *in vivo*. **a**, **b**, *TGFB1*–*TGFB3* (**a**) *or Tgfb1*–*Tgfb3* (**b**) mRNA expression levels in sorted cell populations from disaggregated human CRCs (**a**) or mouse liver metastasis (**b**). First, epithelial cancer cells, immune cells, endothelial cells and CAFs were sorted with labelled antibodies against EPCAM, CD45 ($n = 14$ patients for each), CD31 and FAP ($n = 6$ for each) for human and PDGFRB for mouse. Subsequently, RNA from sorted cells was extracted and analysed by microarray (human) or reverse transcription–quantitative PCR (RT–qPCR) (mouse). **a,** Standardized expression data are shown as *z* score. Individual values and Tukey box plots; *P* values are for *TGFB1*: epithelial versus leukocytes, 7.5×10^{-6}; epithelial versus endothelial, 6.2×10^{-4}; epithelial versus CAFs, 6.2×10^{-4}; and the Kruskal–Wallis (KW) *P* value is 6.3×10^{-6}. For *TGFB2*: CAFs versus epithelial, 6.2×10^{-4}; CAFs versus leukocytes, 8.4×10^{-4}; CAFs versus endothelial, 0.0051; the Kruskal–Wallis *P* value is 0.0017. For *TGFB3*: epithelial versus CAFs, 6.2×10^{-4}; epithelial versus leukocyte, 0.0409; epithelial versus endothelial, 0.0150; leukocyte versus CAFs, 6.2×10^{-4}; endothelial versus CAFs, 0.0051; the Kruskal–Wallis *P* value is 8.5×10^{-5}. Direct comparisons using two-sided Mann–Whitney *U* test. Data are shown as mRNA expression levels normalized to epithelial cells, mean ± s.e.m. from technical triplicates; *Tgfb2* was undetectable (ND) in leukocytes. Note that in both settings, *TGFB1* is expressed by all stromal cell types, *TGFB2* and *TGFB3* are mainly expressed by CAFs. Expression in epithelial cancer cells is comparatively low. **c**–**h**, Immunohistochemistry for TGFβ target gene products in the TME of MTO138-derived liver metastases, representing two independent experiments (**c**, **e**, **g**), as well as for the inhibition of stromal TGFβ signalling by galunisertib (**d**, **f**, **h**). **c**, **d**, Micrographs of sections of liver metastases stained for pSMAD2. **e**, **f**, Staining reveals expression of the TGFβ target CALD1 exclusively in the metastatic TME. Treatment with galunisertib for three days abrogates expression of this cytoplasmic protein. **g**, **h**, Staining for the TGFβ target IGFBP7, which is exclusively expressed in the metastatic TME. Treatment with galunisertib for three days abrogates expression of this secreted protein. Scale bars, 100 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

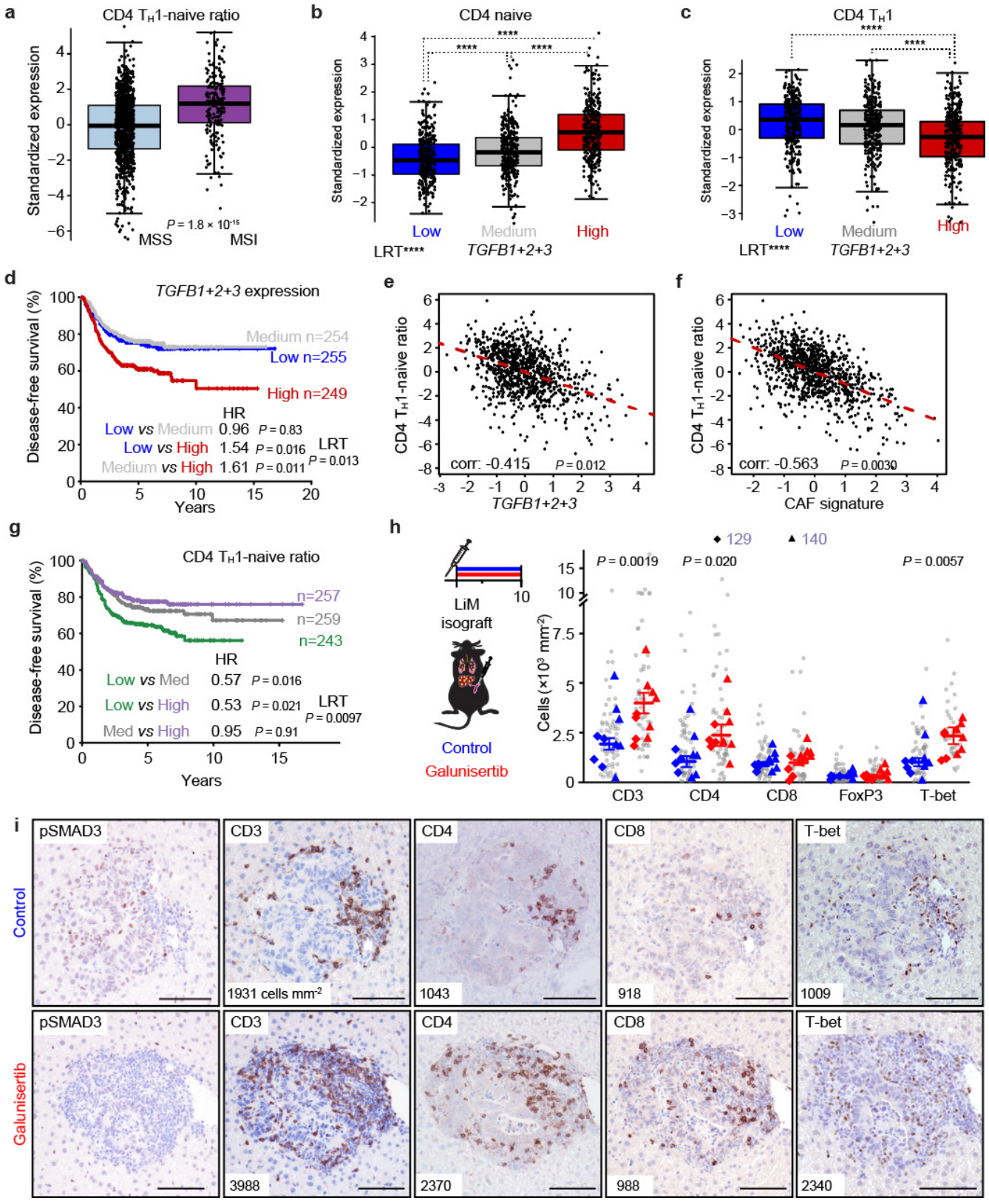


Extended Data Figure 8 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 8 | Characterization of MTOs in metastatic experiments. a, Representative examples of GFP positivity measured by flow cytometry, indicating genetic recombination in the *UBC*^{creERT2};*Tgfbr2*$^{fl/fl}$ genetic model (representative of three independent experiments). **b**, Liver metastases counted five weeks after intrasplenic injection of LAKP MTO54 or MTO220 and treatment from day 2 to day 14 with galunisertib (red; $n = 4$ mice for MTO54 and $n = 10$ for MTO220) or control (blue; $n = 5$ mice for MTO54 and $n = 10$ for MTO220), mean ± s.e.m., 95% confidence interval for MTO220 is (−3–0), from a two-sided Mann–Whitney *U* test. **c**, Stromal TGFβ activation markers in LAKP MTO liver tumours, both supporting the classification of these tumours as CMS4 (Fig. 1i) and explaining the efficacy of galunisertib in **b**; representative of two independent experiments. **d**, Number of nodules five weeks after injection of LAKP empty guide or LAKPS (LAKP + S) MTOs and treatment with galunisertib (red) or control (blue), mean ± s.e.m., 95% confidence interval for LAKPS is (−554 to −176), $n = 5$ mice per condition; two-sided Mann–Whitney *U* test. **e**, Number of liver nodules (mean ± s.e.m.) four weeks after intrasplenic injection of MTOs (either unlabelled or with mCherry–luciferase vector, or after CRISPR knockout of the mCherry–luciferase coding region for MTO138), after treatment. 95% confidence intervals are: for MTO93, (−113−0), $n = 5$ mice per condition; MTO93-LUC: (−104 to −4), $n = 5$ control and $n = 4$ galunisertib; MTO140, (−297 to −21), $n = 7$ control and $n = 5$ galunisertib; MTO140-LUC, (−251 to −21), $n = 5$ per condition; MTO138-LUC, (−86 to −8), $n = 5$ per condition; MTO138-KO-LUC, (−70 to −13), $n = 7$ per condition; two-sided Mann–Whitney *U* test. **f**, Longitudinal intravital bioluminescence imaging (BLI) quantification (photons s^{-1}, normalized to day 0) of an intrasplenic metastasis initiation experiment with MTO140 in C57BL/6J mice treated with galunisertib or control until day 14, $n = 5$ mice per condition. Points and lines represent individual mice, trend lines (bold) show a LOESS model with 95% confidence interval (grey band). 95% confidence interval for the difference at day 25: (−222.7 to −23.9), two-sided Mann–Whitney *U* test. **g**, BLI (as in **f**) of a metastasis-initiation experiment with MTO138 in C57BL/6J mice, $n = 24$ control mice, $n = 15$ galunisertib. **h**, BLI (as in **f**) of galunisertib treatment in intrasplenic liver colonization assays with MTO129 or MTO138, in C57BL/6J ($n = 7$ mice per condition, except for MTO138 + Gal, $n = 11$) or athymic *nu/nu* mice ($n = 5$ per condition). Note the complete absence of treatment effect in the mice lacking functional T cells. **i**, Number of liver nodules (mean ± s.e.m.) in mice intrasplenically injected with MTO129 and treated with galunisertib alone or with antibodies against CD4 or IgG control. 95% confidence interval for galunisertib versus control is (−79 to −8), $n = 5$ mice per condition; two-sided Mann–Whitney *U* test. **j**, Number of liver nodules (mean ± s.e.m.) after intrasplenic injection of MTO93 in syngeneic, *Lgr5*$^{eGFP\text{-}creERT2}$ mice—from the colonies that gave rise to the 3× and 4× genetic models—treated with galunisertib (red) or control (blue). 95% confidence interval is (−13 to −1), $n = 20$ control mice and $n = 25$ for galunisertib; two-sided Mann–Whitney *U* test. Scale bars, 500 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

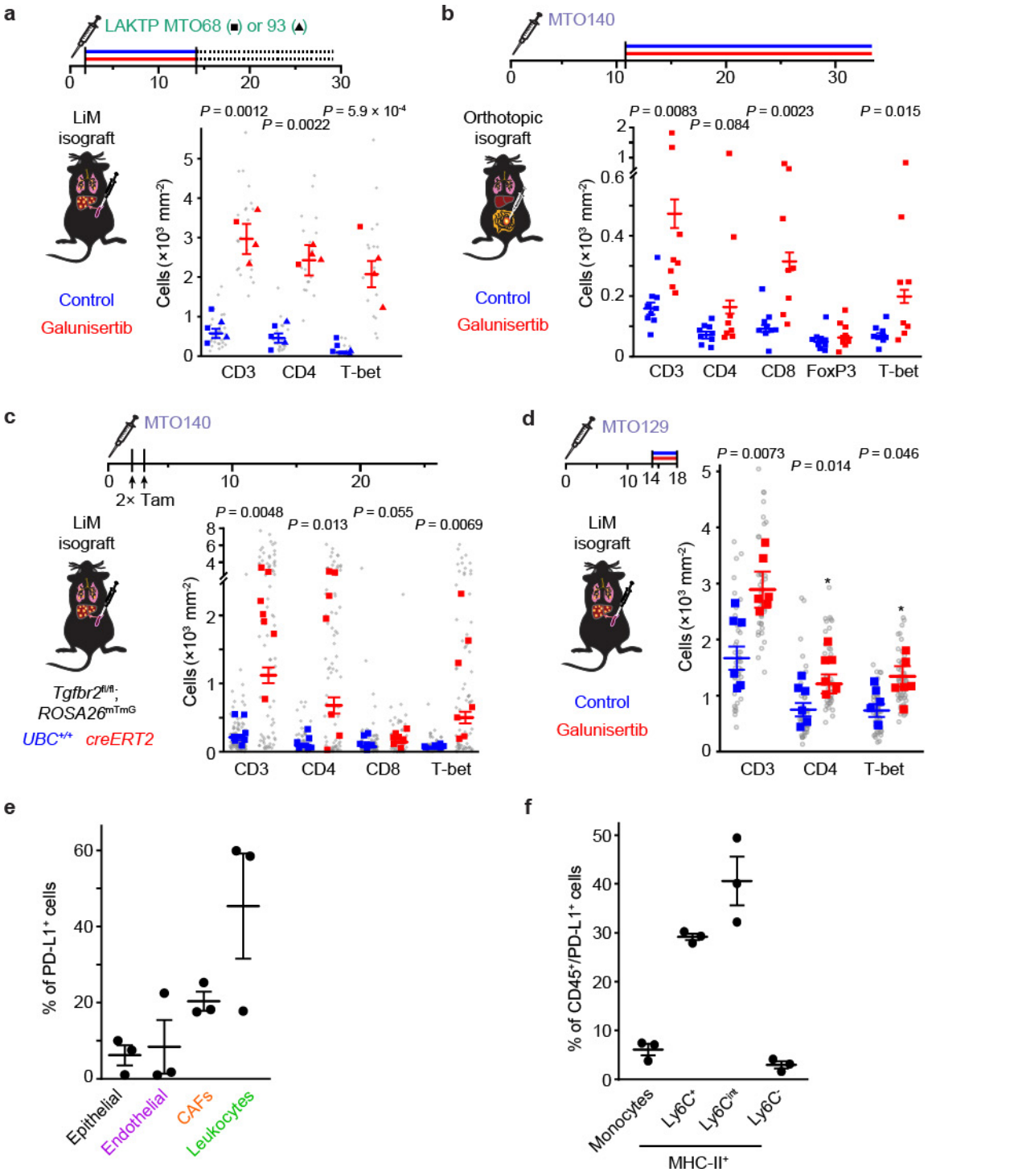


Extended Data Figure 9 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 9 | Immune modulation by TGFβ in CRC. A human meta-cohort was used to analyse the interaction of TGFβ expression levels and T-cell-activation state. **a**, Expression signature ratio between active CD4 T_H1 and naive CD4 T cells differs significantly between MSS and MSI patients. Standardized expression for individual patients and Tukey box plots; MSS, $n = 981$ samples; MSI, $n = 198$; two-sided Wald test. **b**, **c**, Expression signature for naive CD4 T cells (**b**) and T_H1-differentiated CD4 T cells (**c**) in patients with low, medium and high expression of *TGFB1–TGFB3*. Standardized expression for individual patients and Tukey box plots. *P* values (**b**): low versus medium ($n = 327$), 1.7×10^{-4}; low versus high, 2.2×10^{-16}; medium versus high, 2.2×10^{-16}. *P* values (**c**): low versus medium, 0.0503; low versus high, 6.8×10^{-14}; medium versus high, 8.1×10^{-8}; $n = 327$ patient samples per group; two-sided Wald test. The log-likelihood ratio test *P* value is 2.2×10^{-16} for **b** and 2.3×10^{-14} for **c**. **d**, *TGFB1–3* mRNA expression levels predict poor prognosis. Hazard risk (HR) 95% confidence intervals are: low versus medium, (0.65–1.42); low versus high, (1.08–2.20); medium versus high, (1.12–2.32); $n = 255$ patient samples (low), $n = 254$ (medium) and $n = 249$ (high); two-sided Wald test. **e**, **f**, Correlation between ratio of CD4 T_H1 versus naive CD4 signatures and *TGFB1–3* mRNA expression levels (**e**) or expression of a FAP^+ CAF signature (**f**) in $n = 981$ MSS patient samples. Pearson correlation coefficients are indicated; two-sided Wald test. **g**, Kaplan–Meier curves for relapse-free survival for patients according CD4 T_H1-naive signatures ratio. Hazard risk 95% confidence intervals are: low versus medium (0.40–0.81), low versus high (0.37–0.78), medium versus high (0.64–1.41); $n = 243$ (low), $n = 259$ (medium) and $n = 257$ (high); two-sided Wald test. **h**, Cell densities in micrometastases 10 days after injection, treated with vehicle (control, blue) or galunisertib (red); shown are individual MTO129 and MTO140 tumour values and mouse means (diamonds and triangles, respectively), together with group mean ± s.e.m. derived from a mixed-effects linear model; $n = 4$ mice per condition (MTO129) and $n = 6$ for MTO140. **i**, Representative images of immunohistochemistry of the MTO140 micrometastases quantified in **h**. Mean cell densities (cells mm^{-2}) are indicated. Images represent two independent experiments. Scale bars, 100 μm.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | TGFβ inhibition and the tumour immune microenvironment. **a**–**d**, Densities of tumour-infiltrating lymphocytes in galunisertib (Gal, red) or control (blue) treated liver metastases derived from primary LAKTP MTOs (**a**), in treated orthotopically injected MTO-derived primary cancers (**b**), in liver metastases in the $UBC^{creERT2}$;$Tgfbr2^{fl/fl}$ background compared to $UBC^{+/+}$;$Tgfbr2^{fl/fl}$ control (**c**), or two days after start of treatment in established liver metastases (**d**; treatment started on day 14). Shown are individual tumours (grey circles, **a**, **c**, **d**), mouse means (squares, or triangles for MTO93, **a**), and group means ± s.e.m. Data are analysed with a mixed-effects linear model; in (**a**) $n = 5$ (control) or $n = 4$ (galunisertib) mice per condition; in (**b**), $n = 9$ (control) or $n = 8$ (galunisertib) mice per condition; in (**c**), $n = 5$ ($UBC^{+/+}$) or $n = 7$ ($UBC^{creERT2}$;$Tgfbr2^{fl/fl}$) mice per condition; and in (**d**), $n = 6$ mice per condition. **e**, Distribution of cell types within the PD-L1$^+$ population (mean ± s.e.m.) in microdissected tumours (euthanized at day 18, four days after treatment start); $n = 3$ mice. **f**, Myeloid cell types (gated for CD45/(CD11b and/or CD11c)) within the CD45$^+$ PD-L1$^+$ population. Mean ± s.e.m. percentage of $n = 3$ mice.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature25501

TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells

Sanjeev Mariathasan[1]*, Shannon J. Turley[1]*, Dorothee Nickles[1]*, Alessandra Castiglioni[1], Kobe Yuen[1], Yulei Wang[1], Edward E. Kadel III[1], Hartmut Koeppen[1], Jillian L. Astarita[1], Rafael Cubas[1], Suchit Jhunjhunwala[1], Romain Banchereau[1], Yagai Yang[1], Yinghui Guan[1], Cecile Chalouni[1], James Ziai[1], Yasin Şenbabaoğlu[1], Stephen Santoro[1], Daniel Sheinson[1], Jeffrey Hung[1], Jennifer M. Giltnane[1], Andrew A. Pierce[1], Kathryn Mesh[1], Steve Lianoglou[1], Johannes Riegler[1], Richard A. D. Carano[1], Pontus Eriksson[2], Mattias Höglund[2], Loan Somarriba[3], Daniel L. Halligan[3], Michiel S. van der Heijden[4], Yohann Loriot[5], Jonathan E. Rosenberg[6], Lawrence Fong[7], Ira Mellman[1], Daniel S. Chen[1], Marjorie Green[1], Christina Derleth[1], Gregg D. Fine[1], Priti S. Hegde[1], Richard Bourgon[1] & Thomas Powles[8]

Therapeutic antibodies that block the programmed death-1 (PD-1)–programmed death-ligand 1 (PD-L1) pathway can induce robust and durable responses in patients with various cancers, including metastatic urothelial cancer[1–5]. However, these responses only occur in a subset of patients. Elucidating the determinants of response and resistance is key to improving outcomes and developing new treatment strategies. Here we examined tumours from a large cohort of patients with metastatic urothelial cancer who were treated with an anti-PD-L1 agent (atezolizumab) and identified major determinants of clinical outcome. Response to treatment was associated with $CD8^+$ T-effector cell phenotype and, to an even greater extent, high neoantigen or tumour mutation burden. Lack of response was associated with a signature of transforming growth factor β (TGFβ) signalling in fibroblasts. This occurred particularly in patients with tumours, which showed exclusion of $CD8^+$ T cells from the tumour parenchyma that were instead found in the fibroblast- and collagen-rich peritumoural stroma; a common phenotype among patients with metastatic urothelial cancer. Using a mouse model that recapitulates this immune-excluded phenotype, we found that therapeutic co-administration of TGFβ-blocking and anti-PD-L1 antibodies reduced TGFβ signalling in stromal cells, facilitated T-cell penetration into the centre of tumours, and provoked vigorous anti-tumour immunity and tumour regression. Integration of these three independent biological features provides the best basis for understanding patient outcome in this setting and suggests that TGFβ shapes the tumour microenvironment to restrain anti-tumour immunity by restricting T-cell infiltration.

Pre-treatment tumour samples from a large phase 2 trial (IMvigor210) investigating the clinical activity of PD-L1 blockade with atezolizumab in metastatic urothelial cancer (mUC) were used for an integrated evaluation of biomarkers (Extended Data Fig. 1a; Supplementary Discussion). We categorized patients who responded to treatment with a complete or partial response as responders and compared them with non-responders, who displayed stable or progressive disease. As previously reported[2,4], PD-L1 expression on immune cells (more than 5% of immune cells in mUC are detected by the PD-L1 antibody SP142) was significantly associated with response (Fig. 1a). By contrast, PD-L1 expression on tumour cells was not associated with response (Extended Data Fig. 1b). We next performed transcriptome RNA sequencing (RNA-seq) in 298 tissue samples and assessed correlations with PD-L1 expression on immune cells and response. A gene set associated with $CD8^+$ T-effector (T_{eff}) cells[4,5] was highly correlated with PD-L1 expression on immune cells (Extended Data Fig. 1c, d; Supplementary Discussion). This gene set was also significantly associated with response, particularly with complete response, and with overall survival (Fig. 1b, c).

mUC is characterized by one of the highest somatic mutation rates[6,7]. In mUC, tumour mutation burden (TMB) correlates with response to immune checkpoint inhibitors[4,5]. We confirmed these findings (Fig. 1d, e) and showed that computationally predicted tumour neoantigen burden behaved similarly (Extended Data Fig. 1e, f), suggesting that the TMB reflects an increased potential for immunogenicity[8–11]. We next assessed the transcriptional and mutational correlates of TMB in mUC. The pathways that were most significantly associated with TMB were those involved in the cell cycle, DNA replication and DNA damage response (DDR; Extended Data Fig. 1g; Supplementary Table 1). Signatures for these pathways were correlated with *MKI67* and thus with proliferation (Extended Data Fig. 1h). Expression levels of *APOBEC3A* and *APOBEC3B*, which encode two cytidine deaminases that are upregulated in urothelial and other cancers[12,13], were also correlated with TMB and response (Extended Data Fig. 1i, j; Supplementary Tables 2, 3). Finally, tumours with one or more mutations in DDR or cell cycle regulator gene sets had significantly higher TMB and response rates (Fig. 1f, g; Extended Data Fig. 1k, l; Supplementary Tables 4, 5).

Next, we investigated features beyond $CD8^+$ T-cell immunity and TMB that we associated with response. Gene set enrichment analysis identified the cytokine–cytokine-receptor gene set as the only feature associated with non-response (Extended Data Fig. 2a; Supplementary Table 6). The most significantly associated genes within this pathway were *IFNGR1* and genes involved in the TGFβ signalling pathway: *TGFB1*, *ACVR1* and *TGFBR2* (Supplementary Table 3). Whereas IFNγ is known to have favourable effects on anti-tumour immunity, persistent signalling by this cytokine has been associated with adaptive resistance to checkpoint therapy[14–17]. We observed significantly enhanced expression of *IFNG* in responders, whereas *IFNGR1* expression was significantly higher in non-responders (Extended Data Fig. 2b, c). TGFβ is a pleiotropic cytokine that is associated with poor prognosis in multiple tumour types[18–20] and is thought to have a protumorigenic role in advanced cancers by promoting immunosuppression, angiogenesis, metastasis, tumour cell epithelial-to-mesenchymal transition (EMT), fibroblast activation and desmoplasia[19,21–23]. The two top-scoring TGFβ pathway genes in our data represent a TGFβ

[1]Genentech, South San Francisco, California 94080, USA. [2]Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, Lund, Skåne 223 81, Sweden. [3]Fios Genomics, Edinburgh EH16 4UX, UK. [4]Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands. [5]Department of Cancer Medicine, Institut Gustave Roussy, University of Paris Sud, 94800 Villejuif, France. [6]Genitourinary Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [7]University of California San Francisco, Helen Diller Family Comprehensive Cancer Center, San Francisco, California 94158, USA. [8]Barts Experimental Cancer Medicine Centre, Barts Cancer Institute, Queen Mary University of London, London EC1M 6BQ, UK.
*These authors contributed equally to this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

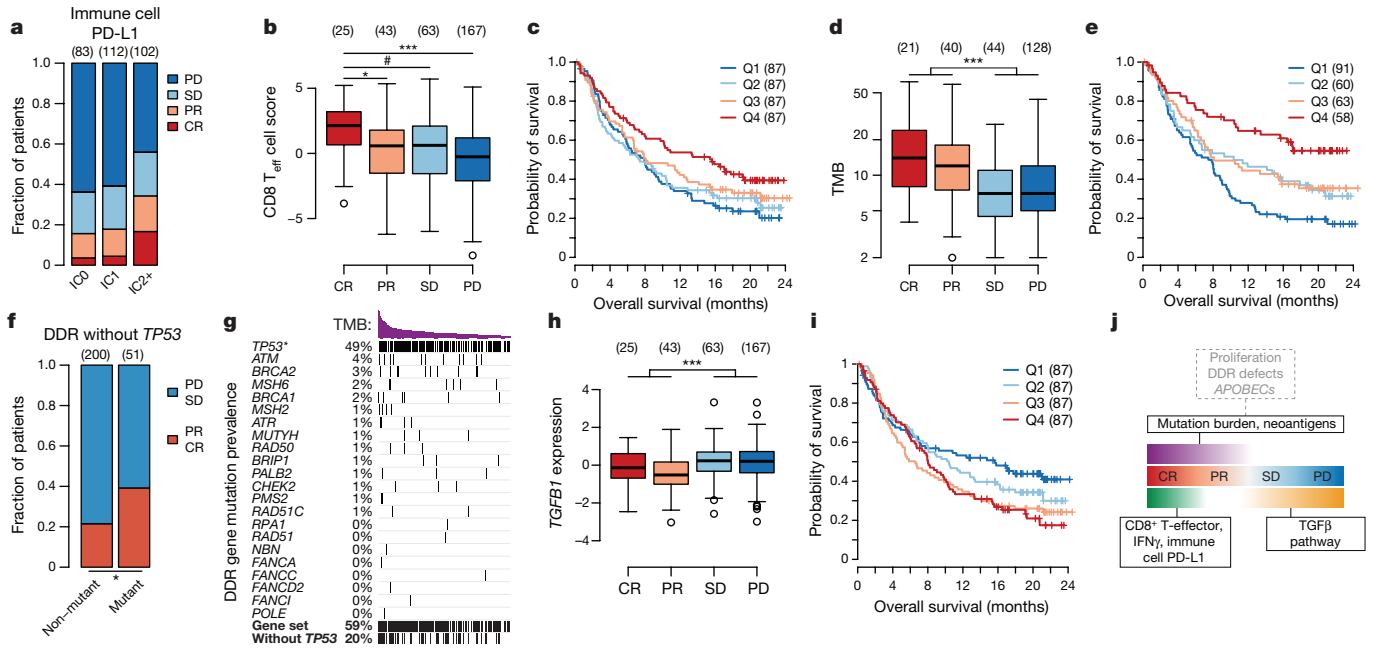


Figure 1 | Three core biological pathways are associated with response to atezolizumab. a, PD-L1 on immune cells is associated with response (two-sided Fisher's exact test $P = 0.0038$). IC2/3 tumours had a significantly higher complete response (CR) rate ($P = 0.0006$). PD, progressive disease; SD, stable disease; PR, partial response. PD-L1 expression on immune cells (IC) is assessed by SP142 immunohistochemistry assay and scored as IC0 (<1%), IC1 (≥1% and <5%), or IC2/3 (≥5%). **b**, $CD8^+$ T_{eff}-signature score is positively associated with response (two-tailed *t*-test, $P = 0.0087$), with association driven by the complete response group (CR versus PR, $P = 0.0388$; CR versus SD, $P = 0.0668$; CR versus PD, $P = 0.0003$). **c**, $CD8^+$ T_{eff}-signature quartiles are significantly associated with overall survival (likelihood ratio test, $P = 0.0092$). **d**, **e**, TMB is positively associated with response to atezolizumab (two-tailed *t*-test, $P = 6.9 \times 10^{-7}$) and overall survival (likelihood ratio test, $P = 2.0 \times 10^{-5}$). A similar plot for tumour neoantigen burden is shown in Extended Data Fig. 1e, f. **f**, **g**, There is a significant association between DDR mutation status and response (**f**; two-sided Fisher's exact test, $P = 0.0117$ excluding *TP53*) and TMB (**g**), both with (two-sided Fisher's exact test, $P = 6.01 \times 10^{-8}$) and without inclusion of *TP53* ($P = 1.95 \times 10^{-5}$). **h**, **i**, *TGFB1* gene expression is significantly associated with non-response (two-tailed *t*-test, $P = 0.00011$) and reduced overall survival (likelihood ratio test, $P = 0.0096$). **j**, The relationship between response and three core biological pathways. $\#P < 0.10$, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Sample sizes are given in parentheses. Q1–Q4, quartiles, lowest to highest. Box plots show median, whiskers represent minimum and maximum excluding outliers, and circles are outliers.

ligand, *TGFB1*, and a TGFβ receptor, *TGFBR2*. Both showed increased mean expression in non-responders and association with reduced overall survival (Fig. 1h, i; Extended Data Fig. 2d, e). Taken together, these results suggest that the impact of checkpoint inhibition on patient outcome in mUC is dictated by three core biological pathways: (i) pre-existing T-cell immunity and (ii) TMB are positively associated with outcome, whereas (iii) TGFβ is associated with lack of response and reduced survival (Fig. 1j).

Most human solid tumours exhibit one of three distinct immunological phenotypes: immune inflamed, immune excluded, or immune desert[1,24]. Previous evidence, mostly from melanomas, has suggested that inflamed tumours are the most responsive to checkpoint blockade[24,25], but the relevance of the tumour–immune phenotype to the response of mUCs was previously unknown. In our mUC cohort, a large proportion of tumours (47%) exhibited the excluded phenotype; by contrast, only 26% and 27% of tumours exhibited the inflamed and desert phenotypes, respectively (Fig. 2a, b; Extended Data Fig. 2f). The mean signal for the $CD8^+$ T_{eff} signature was lowest in the desert phenotype, intermediate in the excluded phenotype and highest in inflamed tumours (Fig. 2c), and was significantly associated with response only in inflamed tumours, consistent with a model in which the absence of $CD8^+$ T cells, or their spatial separation from tumour cells, makes the signature irrelevant.

The observed proximity of $CD8^+$ T cells to desmoplastic stroma in immune-excluded tumours (Fig. 2a; Extended Data Fig. 2f) suggests that the relevance of TGFβ in this phenotype may be driven by its impact on stromal cells. To measure TGFβ pathway activity specifically in fibroblasts, we generated a pan-fibroblast TGFβ response signature (F-TBRS). Average expression of this signature was low in immune deserts but significantly higher in inflamed and excluded tumours. Consistent with a role for TGFβ pathway activation in fibroblasts in the tumour microenvironment (TME), the F-TBRS was significantly associated with non-response only in excluded tumours (Fig. 2d). TMB was significantly associated with response in both excluded and inflamed tumours (Fig. 2e).

To better understand how the three core pathways relate to one another and to measure the relative strengths of their association with outcome, we performed a statistical analysis of competing models. Logistic regression pseudo-R^2 was used as a measure of 'explained variance' in patient response[26].

In immune desert tumours, the pathways showed negligible explanatory power (Fig. 2f). For excluded tumours, both the F-TBRS signature and TMB were significantly associated with response, and combining the two provided a significant improvement over either term in isolation. By contrast, in inflamed tumours, the $CD8^+$ T_{eff} signature combined with TMB exhibited the strongest correlation with response. In an analysis in which all the samples were combined, a model incorporating the three core pathways significantly improved on all single- and two-pathway models (Fig. 2f; Extended Data Fig. 2g), demonstrating that the information provided by each pathway is at least partially independent.

The Cancer Genome Atlas (TCGA) molecular-subtype taxonomy[12] has been widely investigated with respect to mUC, but with inconsistent results. Here we contrast the TCGA taxonomy with the Lund taxonomy, which includes a genomically unstable subtype[27,28] (Extended Data Fig. 3). Consistent with the importance of TMB, we observed that the genomically unstable subtype was significantly enriched for response (Fig. 3). This effect could not, however, be attributed to TMB alone, as the TCGA luminal II subtype was similarly enriched for high-TMB tumours (Extended Data Fig. 4a, b). Instead, we identified the source

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

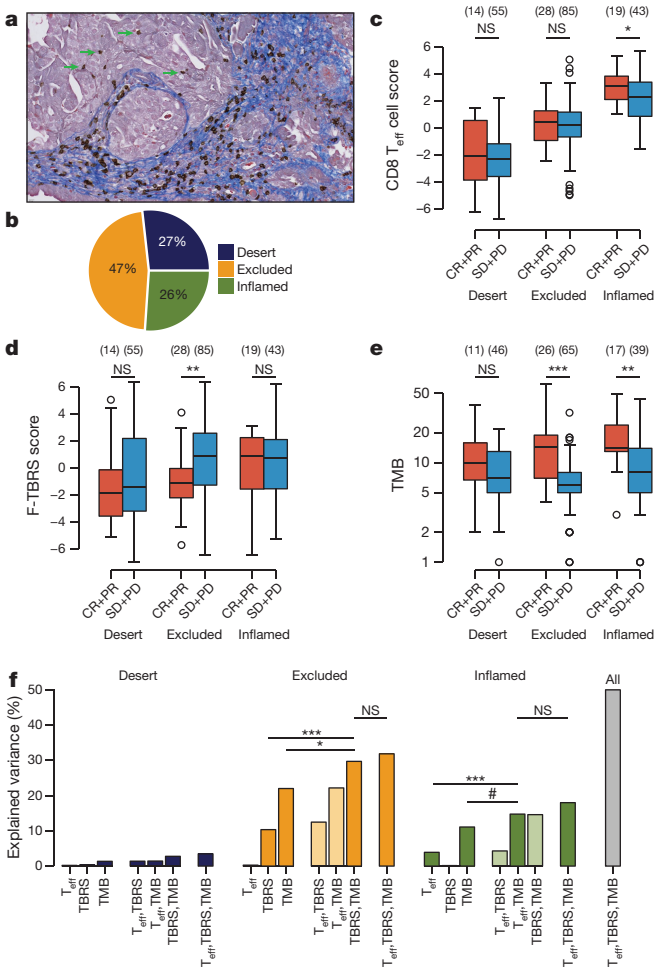


Figure 2 | TGFβ is associated with lack of response in the excluded tumour–immune phenotype. a, Combined CD8 immunohistochemistry–trichrome stain. $CD8^+$ T cells (stained with 3,3′-diaminobenzidine (DAB), brown) are primarily within collagenous stroma (blue). Rare $CD8^+$ T cells are interspersed between tumour cells (green arrows). **b**, Distribution of tumour–immune phenotypes in IMvigor210. **c**, $CD8^+$ T_{eff} score is significantly associated with response only in the inflamed phenotype (two-tailed *t*-test, $P = 0.042$). **d**, F-TBRS is significantly associated with response only in the excluded phenotype (two-tailed *t*-test, $P = 0.0066$). **e**, TMB is significantly associated with response in the excluded and inflamed phenotypes (two-tailed *t*-test, $P = 0.00009$ and 0.00258, respectively). **f**, Explained variance in patient response for generalized linear models fit using single core pathways or pathway combinations. In immune desert tumours ($n = 57$), core pathways showed negligible explanatory power. For excluded tumours ($n = 91$), F-TBRS and TMB were associated with response as single terms; combining the two provided statistically significant improvement over single terms (likelihood ratio, $P = 6.23 \times 10^{-5}$ and 0.02125, respectively). For inflamed tumours ($n = 56$), $CD8^+$ T_{eff} and TMB were associated with response as single terms; combining the two provided statistically significant improvement over single terms (likelihood ratio, $P = 0.00099$ and 0.0557, respectively). Standard forward selection applied to all patients (grey) identified a three-pathway model as a significantly better fit than all single- or two-pathway models (Extended Data Fig. 2g). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. NS, nonsignificant. For box plots, centre mark is median, whiskers are minimum/maximum excluding outliers, and circles are outliers. Sample sizes are given in parentheses.

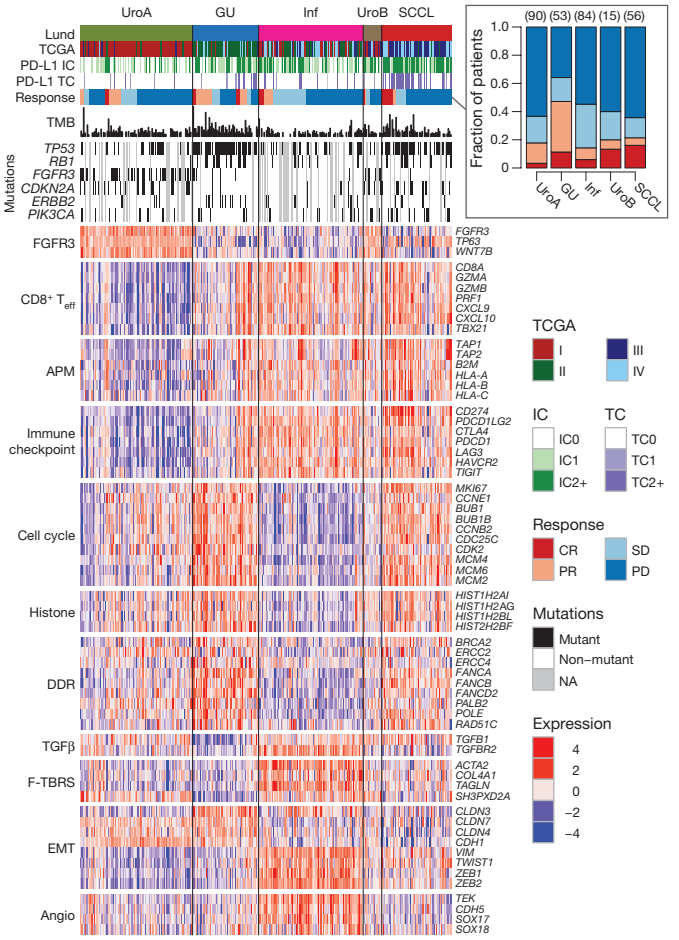


Figure 3 | Relationship between molecular subtypes (Lund scheme) and core biological pathways. Rows of the heat map show gene expression (*Z* scores) grouped by pathway. Top right, response versus Lund molecular subtype, showing that the genomically unstable (GU) subtype had a significantly higher response rate (two-sided Fisher's exact test, $P = 1.6 \times 10^{-5}$). Sample sizes are given in parentheses. Inf, infiltrated; SCCL, basal/SCC-like; TC, tumour cells; UroA, urothelial-like A; UroB, urothelial-like B.

of the difference by separately evaluating patients classified as TCGA luminal II only, Lund genomically unstable only, or both (Extended Data Fig. 4c). Among these groups, genomically unstable-only tumours had the lowest $CD8^+$ T_{eff} signature scores and lowest TMB, but responded favourably; by contrast, luminal II-only tumours showed sharply increased F-TBRS scores and responded poorly (Extended Data Fig. 4d, e). These results demonstrate the importance of interactions among the three core pathways in determining response. The Lund subtypes are discussed further in the Supplementary Discussion.

To test our hypothesis that physical exclusion of T cells by the stromal barrier limits response to atezolizumab in immune-excluded tumours, we studied the EMT6 mouse mammary carcinoma model to determine whether there is a role for TGFβ-activated stroma in this context. The EMT6 model exhibits the immune-excluded phenotype (Extended Data Fig. 5a–d) and expresses all TGFβ isoforms as well as PD-L1 (Extended Data Fig. 5e, f). Although therapeutic blockade of PD-L1 or TGFβ alone had little or no effect, mice treated with antibodies against both PD-L1 and TGFβ exhibited a significant reduction in tumour burden (Fig. 4a, b; Extended Data Fig. 5g, h). Regression of EMT6 tumours in these studies was completely dependent on $CD8^+$ T cells (Extended Data Fig. 5i). These findings were reproduced in a second relevant mouse tumour model, MC38 (Extended Data Fig. 5j–n).

Combined antibody blockade also led to a significant increase in tumour-infiltrating T cells (Extended Data Fig. 6a), particularly $CD8^+$ T_{eff} cells (Fig. 4c, d; Extended Data Fig. 6b). Blockade of TGFβ, alone or in combination with anti-PD-L1, had no effect on the number of $CD4^+$ T-regulatory (T_{reg}) cells in the tumour (Extended Data Fig. 6c–e). RNA-seq data revealed that the $CD8^+$ T_{eff} signature was increased in mouse tumours treated with a combination of anti-PD-L1 and anti-TGFβ (Fig. 4e). Quantitative histopathology demonstrated that T cell

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

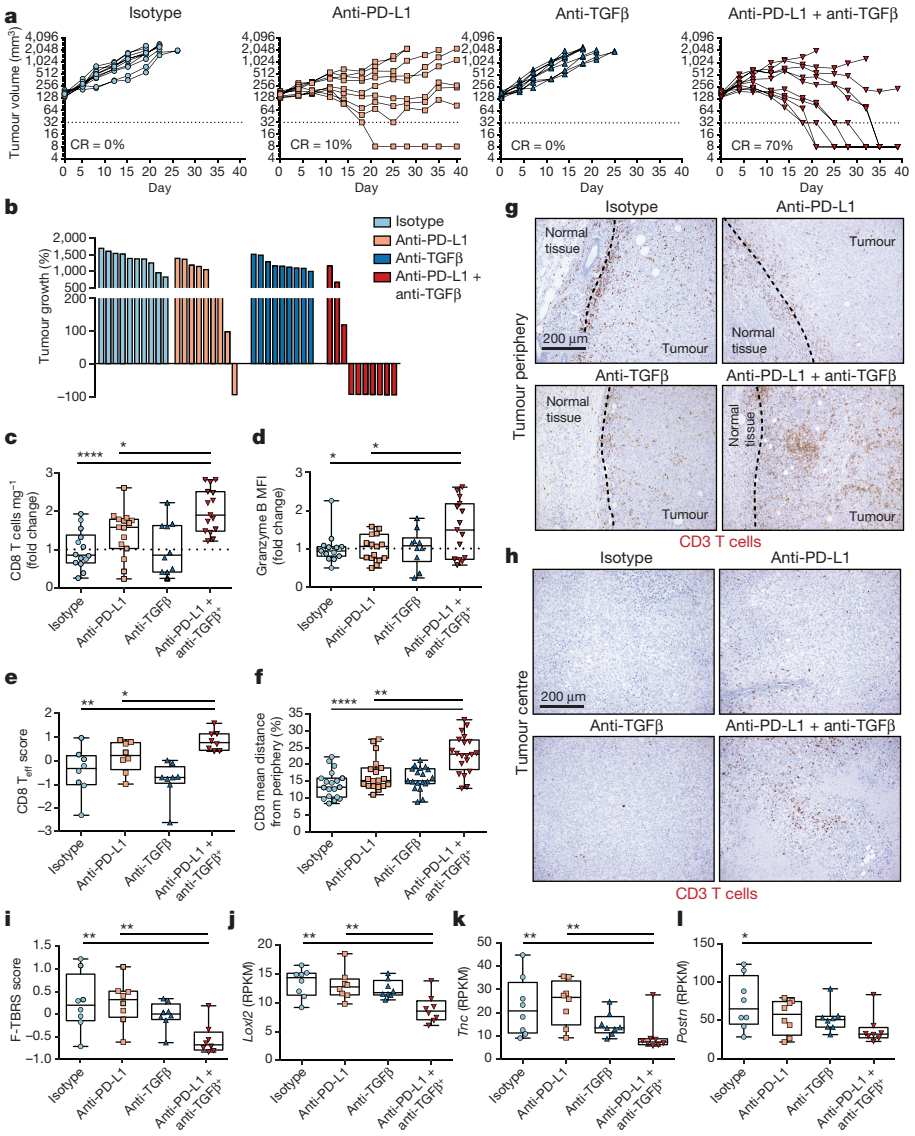


Figure 4 | Tumour regression and changes in TME following therapeutic anti-TGFβ and anti-PD-L1 treatment in EMT6 tumours. **a**, Tumour growth curves. **b**, Change in tumour volume compared to baseline. Six independent experiments, 10 mice per group. **c**, **d**, Fold change relative to the isotype mean in total CD8 T-cell abundance (cells mg^{-1}, **c**) and $CD8^+$ T-cell GRZB mean fluorescence intensity (MFI, **d**) by flow cytometry. Three independent experiments, $n = 15$ mice for all groups except anti-TGFβ, where $n = 10$ mice. **e**, $CD8^+$ T_{eff} signature. One experiment, $n = 8$ mice per group. **f**, Quantification of tumour-infiltrating lymphocyte (TIL) localization by immunohistochemistry. Three independent experiments, $n = 19$ mice for all groups except anti-PD-L1/anti-TGFβ, where $n = 20$ mice; Tukey's honest significant difference multiple-comparison adjustment. **g**, **h**, Representative CD3 staining of tumour periphery (**g**) and centre (**h**). Scale bar, 200 μm. **i**–**l**, F-TBRS signature (**i**) and expression of fibroblast genes (**j**–**l**). RPKM, reads per kilobase of transcript per million mapped reads. One experiment, $n = 8$ mice per group. All data shown in **c**–**l** are from tumours collected at day seven after treatment initiation. For box plots, centre mark is median, and whiskers are minimum/maximum. All *P* values are based on two-sided Mann–Whitney *U*-tests unless otherwise indicated. $*P < 0.05$, $**P < 0.01$, $****P < 0.0001$.

distribution significantly changed following combination therapy, with mean distance from the stromal border increasing, and mean distance from the tumour centre decreasing. However, T-cell localization did not change with either single antibody treatment (Fig. 4f–h; Extended Data Fig. 6f). Together, these results suggest that TGFβ inhibition potentiated the ability of anti-PD-L1 to enhance anti-tumour immunity, resulting in optimal T cell positioning and ensuing tumour regression.

Anti-TGFβ treatment significantly reduced TGFβ receptor signalling (as shown by the reduction of phosphorylated (p)SMAD2/3) in EMT6 tumours, particularly in non-immune cells (Extended Data Fig. 6g, h). Given that TGFβ is associated with fibroblast differentiation and EMT[22], we investigated whether the benefits of dual-antibody blockade could be attributed to direct effects on tumour cells or effects on stromal compartments. While single-agent inhibition of TGFβ reduced one of the three EMT signatures that we considered, dual-antibody treatment had no significant impact (Extended Data Fig. 6i). By contrast, the F-TBRS score and expression of canonical fibroblast genes associated with matrix remodelling were significantly reduced in the combined treatment (Fig. 4i–l). Consistent with the phospho-flow analysis that showed no change in pSMAD2/3 in haematopoietic cells, no reduction was observed in two alternate TBRS signatures associated with T cells or macrophages[20] (Extended Data Fig. 6j, k). Blockade of TGFβ can thus synergize with anti-PD-L1 in the EMT6 model to reprogram peritumoural stromal fibroblasts and increase $CD8^+$ T_{eff} cell counts in the tumour bed, leading to robust anti-tumour immunity. Interestingly,

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

although TGFβ inhibition might also be expected to reduce the T_{reg} cell population, we did not observe lower T_{reg} numbers as a result of TGFβ blockade, which therefore did not appear to contribute to the efficacy of the combined treatment.

The comprehensive evaluation of molecular, cellular and genetic factors associated with response and resistance to checkpoint blockade (atezolizumab) in this large cohort of mUC patients has yielded several important conclusions. Three non-redundant factors were found to contribute: (i) pre-existing immunity, as represented by PD-L1 gene expression on immune cells, IFNγ expression and histological correlates of $CD8^+$ T_{eff} activity; (ii) TMB, measured directly (Extended Data Fig. 7), but also reflected in signatures of proliferation and DNA damage response; and (iii) TGFβ-pathway signalling, reflected by a distinct gene expression signature and by pSMAD2/3. These tightly linked factors and their relationships with each other may partly explain why predicting outcomes from PD-L1 expression alone is challenging. The enrichment of the fibroblast TGFβ-response signature in non-responding immune-excluded tumours, combined with results from Lund molecular subtyping and preclinical models showing that co-inhibition of TGFβ and PD-L1 converted tumours from an excluded to an inflamed phenotype, support a model in which TGFβ signalling may counteract anti-tumour immunity by restricting the movement of T-cells in the TME. The observed multifactorial basis of response to immunotherapy may be applicable to other tumour types beyond mUC. Work in this area, across multiple tumour types and therapies, is still in its infancy, but these results open new avenues for disease-agnostic exploration of the mechanisms underlying responses to, and primary immune escape from, cancer immunotherapy.

Code Availability The source code and processed data used for all analyses presented here have been made available in IMvigor210CoreBiologies, a fully documented software and data package for the R statistical computing environment[29]. This package is freely available under the Creative Commons 3.0 license and can be downloaded from http://research-pub.gene.com/IMvigor210CoreBiologies.

Data Availability All raw sequencing data required for RNA-seq analyses have been deposited to the European Genome-Phenome Archive under accession number EGAS00001002556.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 June 2017; accepted 8 January 2018.
Published online 14 February 2018.

1. Herbst, R. S. *et al.* Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* **515,** 563–567 (2014).
2. Powles, T. *et al.* MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature* **515,** 558–562 (2014).
3. Bellmunt, J. *et al.* Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *N. Engl. J. Med.* **376,** 1015–1026 (2017).
4. Rosenberg, J. E. *et al.* Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* **387,** 1909–1920 (2016).
5. Balar, A. V. *et al.* Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. *Lancet* **389,** 67–76 (2017).
6. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).
7. Chalmers, Z. R. *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* **9,** 34 (2017).
8. Snyder, A. *et al.* Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371,** 2189–2199 (2014).
9. Rizvi, N. A. *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348,** 124–128 (2015).
10. Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350,** 207–211 (2015).
11. Chen, D. S. & Mellman, I. Oncology meets immunology: the cancer–immunity cycle. *Immunity* **39,** 1–10 (2013).
12. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507,** 315–322 (2014).
13. Burns, M. B., Temiz, N. A. & Harris, R. S. Evidence for *APOBEC3B* mutagenesis in multiple human cancers. *Nat. Genet.* **45,** 977–983 (2013).
14. Gao, J. *et al.* Loss of IFN-γ pathway genes in tumor cells as a mechanism of resistance to anti-CTLA-4 therapy. *Cell* **167,** 397–404 (2016).
15. Zaretsky, J. M. *et al.* Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N. Engl. J. Med.* **375,** 819–829 (2016).
16. Ribas, A. Adaptive immune resistance: how cancer protects from immune attack. *Cancer Discov.* **5,** 915–919 (2015).
17. Benci, J. L. *et al.* Tumor interferon signaling regulates a multigenic resistance program to immune checkpoint blockade. *Cell* **167,** 1540–1554 (2016).
18. Lin, R.-L. & Zhao, L.-J. Mechanistic basis and clinical relevance of the role of transforming growth factor-β in cancer. *Cancer Biol. Med.* **12,** 385–393 (2015).
19. Massagué, J. TGFβ in cancer. *Cell* **134,** 215–230 (2008).
20. Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* **47,** 320–329 (2015).
21. Derynck, R. & Zhang, Y. E. SMAD-dependent and SMAD-independent pathways in TGF-β family signalling. *Nature* **425,** 577–584 (2003).
22. Morikawa, M., Derynck, R. & Miyazono, K. TGF-β and the TGF-β family: context-dependent roles in cell and tissue physiology. *Cold Spring Harb. Perspect. Biol.* **8,** a021873 (2016).
23. Flavell, R. A., Sanjabi, S., Wrzesinski, S. H. & Licona-Limón, P. The polarization of immune cells in the tumour environment by TGFβ. *Nat. Rev. Immunol.* **10,** 554–567 (2010).
24. Hegde, P. S., Karanikas, V. & Evers, S. The where, the when, and the how of immune monitoring for cancer immunotherapies in the era of checkpoint inhibition. *Clin. Cancer Res.* **22,** 1865–1874 (2016).
25. Tumeh, P. C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515,** 568–571 (2014).
26. Dobson, A. J. & Barnett, A. G. *An* Introduction to Generalized Linear Models 3rd edn. (Chapman & Hall/CRC, 2008).
27. Sjödahl, G. *et al.* A molecular taxonomy for urothelial carcinoma. *Clin. Cancer Res.* **18,** 3377–3386 (2012).
28. Sjödahl, G. *et al.* Toward a molecular pathologic classification of urothelial carcinoma. *Am. J. Pathol.* **183,** 681–691 (2013).
29. R Core Team. *R:* a Language and Environment for Statistical Computing http://www.r-project.org (R Foundation for Statistical Computing, Vienna, Austria, 2016).
30. Rusinova, I. *et al.* Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res.* **41,** D1040–D1046 (2013).
31. Spranger, S., Bao, R. & Gajewski, T. F. Melanoma-intrinsic β-catenin signalling prevents anti-tumour immunity. *Nature* **523,** 231–235 (2015).
32. Damrauer, J. S. *et al.* Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc. Natl Acad. Sci. USA* **111,** 3110–3115 (2014).
33. Hugo, W. *et al.* Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* **165,** 35–44 (2016).
34. Hedegaard, J. *et al.* Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. *Cancer Cell* **30,** 27–42 (2016).

Supplementary Information is available in the online version of the paper.

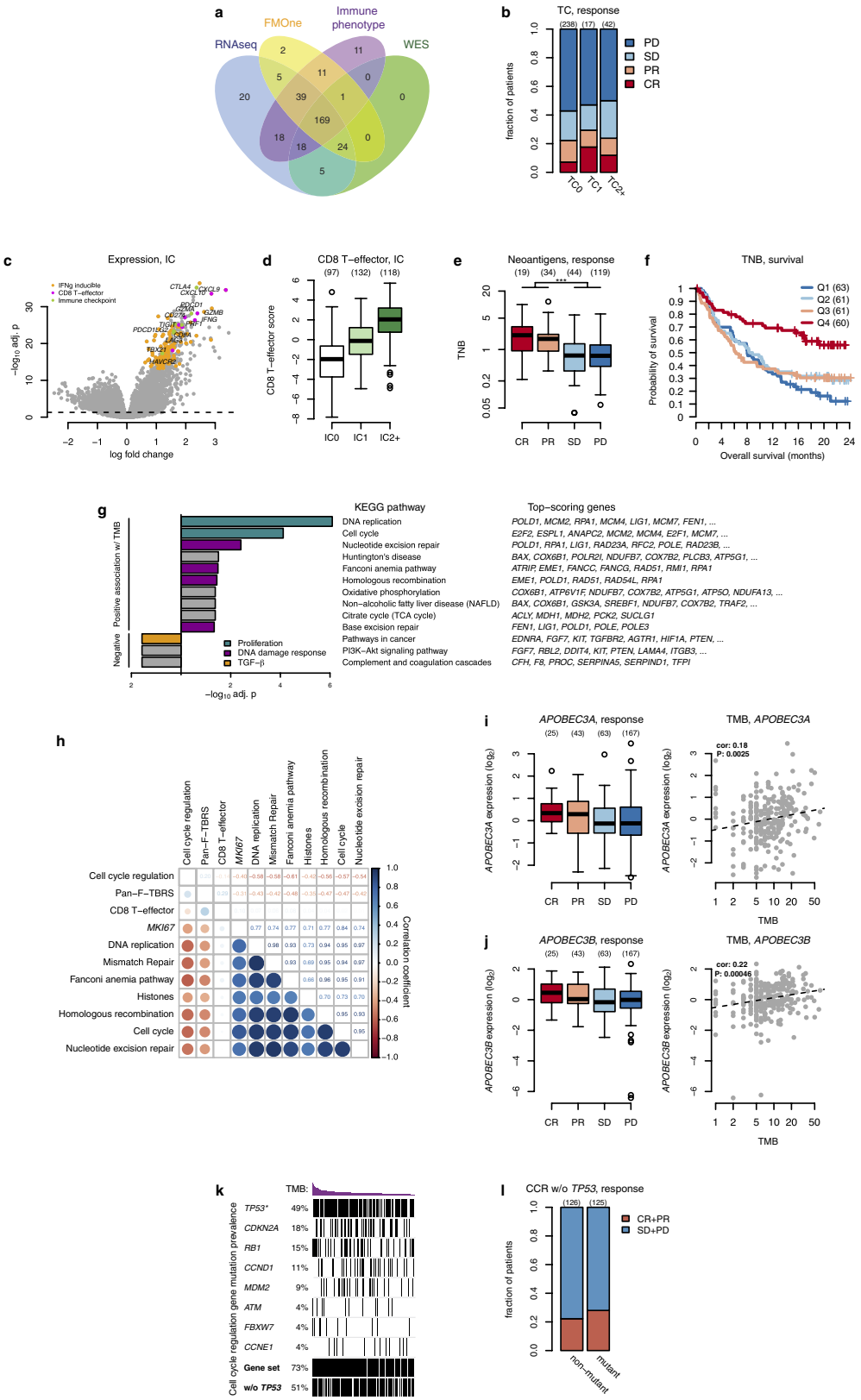
Acknowledgements We thank the patients and their families; all of the investigators and their staff involved in IMvigor210 study; C. Ahearn, S. Lau, C. Havnar, Z. Boyd, S. Sampath, D. Wilson, J. Doss and medical writers at Health Interactions. J.E.R. acknowledges support from P30 CA008748. L.F. acknowledges support from NCI 1R01CA194511.

Author Contributions S.M., S.J.T., D.N., L.F., I.M., D.S.C., J.E.R., Y.L., M.S.H., M.G., C.D., G.D.F., P.S.H., R.Bo and T.P. contributed to the overall study design. S.M., S.J.T., D.N., Y.W., E.E.K., K.Y., R.Ba., Y.G., Y.Ş., S.J., S.L., P.E., M.H., L.S., D.L.H., P.S.H. and R.Bo. performed the biomarker and statistical analyses. H.K., C.C., J.Z., S.S., D.S., J.H., J.M.G., A.A.P. and K.M. conducted microscopy studies. S.J.T., A.C., J.L.A. and R.C. designed all the preclinical experiments and S.J.T., A.C., J.L.A., R.C., Y.Y., C.C., J.Z., Y.Ş., S.S., D.S., J.H., J.M.G., A.A.P., K.M., J.R. and R.A.D.C. analysed the corresponding preclinical data. S.M., S.J.T., D.N., I.M., P.S.H., R.Bo. and T.P. wrote the paper. All authors contributed to data interpretation, discussion of results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to S.M. (mariathasan.sanjeev@gene.com); S.T. (turley.shannon@gene.com) or R.Bo. (bourgon.richard@gene.com).

Reviewer Information *Nature* thanks D. McConkey and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

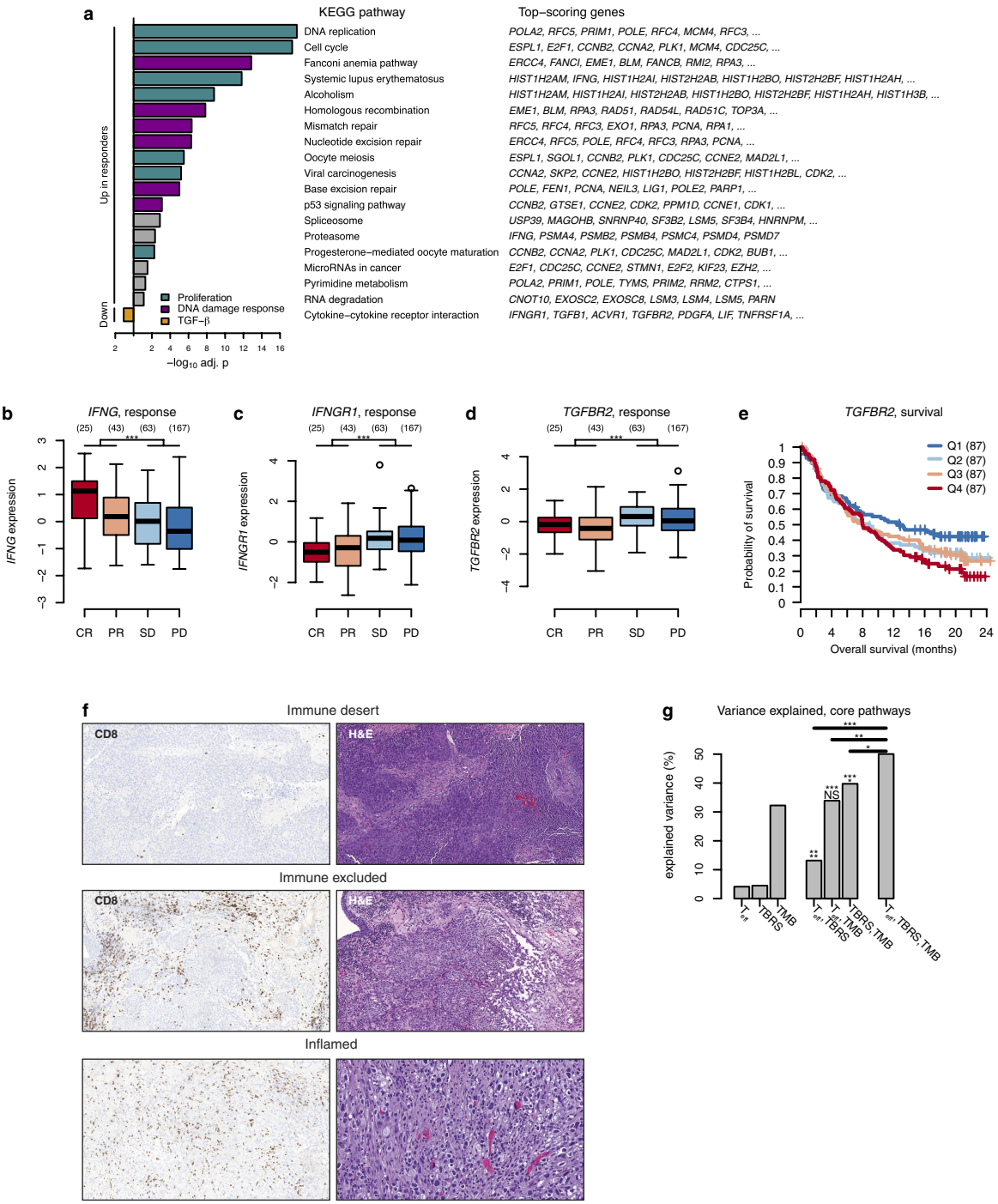


Extended Data Figure 1 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 1 | Molecular correlates of outcome and TMB. **a**, Overlap of the efficacy-evaluable patient populations with assays used in this study ($n = 326$ for one or more of these assays). For gene expression analyses with respect to response, the complete RNA-seq dataset was used ($n = 298$). For gene expression analyses in the context of TMB or immune phenotype, the intersect between RNA-seq and FoundationOne (FMOne, $n = 237$) or immune phenotype ($n = 244$) were used, respectively. For mutation analysis around immune phenotypes, the intersect between FoundationOne and immune phenotype was used ($n = 220$). For associations between response or genes mutation status with TMB, the complete FoundationOne dataset was used ($n = 251$). **b**, PD-L1 protein expression on tumour cells (TC), in contrast to expression on immune cells (Fig. 1a), was not associated with response to atezolizumab (two-sided Fisher's exact test, $P = 0.72$). **c**, **d**, Transcriptional correlates of PD-L1 protein expression on immune cells. **c**, Genes associated with PD-L1 immunohistochemistry positivity on immune cells. Normalized $\log_2$-transformed gene expression was compared with PD-L1 protein expression on immune cells. Interferon-stimulated genes[30] and previously reported CD8 T_{eff} and immune checkpoint-molecule gene sets[4,5] were among the most upregulated (complete list of associated genes is given in Supplementary Table 10). **d**, Association between CD8 T_{eff}-signature score and PD-L1-staining on tumour-infiltrating immune cells. There is a significant positive relationship between the signature score and PD-L1 staining on immune cells (likelihood ratio test, $P = 4.2 \times 10^{-35}$). **e**, **f**, Tumour neoantigen burden (TNB) is associated with outcome. **e**, Box plots showing the relationship between response status and TNB. Shown are the number of mutations based on whole-exome sequencing, filtering for those mutations that are predicted to be expressed neoantigens. TNB is positively associated with response to atezolizumab (two-tailed t-test, $P = 2.7 \times 10^{-9}$). **f**, TNB, split into quartiles, is also associated with overall survival (likelihood ratio test, $P = 0.00015$). **g**, KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways enriched in genes whose expression is correlated with TMB. Shown are adjusted $-\log_{10} P$ values for enrichment of KEGG gene sets that are significantly (false discovery rate < 0.05) enriched in genes that are correlated with TMB (272 samples analysed). Gene sets that are inferred to reflect key underlying biological processes are coloured. Only the top seven genes per set (ranked by single-gene P value) are shown. *ATP5G1* is also known as *ATP5MC1*, *ATP5O* is also known as *ATP5PO*. **h**, Relationship between different gene-expression signatures as well as the single-gene expression values for *MKI67*, a marker for proliferation. Dot size and colour show correlation between signature scores and gene expression (348 samples analysed). Numbers are the Pearson correlation coefficients. Gene-set membership is shown in Supplementary Table 8. **i**, **j**, *APOBEC3A* and *APOBEC3B* gene expression and their association with response and TMB. Both *APOBEC3A* (two-tailed t-test, $P = 0.015$; **i**) and *APOBEC3B* (two-tailed t-test, $P = 0.0025$; **j**) exhibit higher mean expression in responders. For TMB, Pearson correlation coefficients and P values are given. In **j**, the two extreme expression outliers were excluded when calculating correlation between gene expression and TMB. **k**, Mutations in cell-cycle-regulator genes are associated with TMB. Rows represent genes and columns represent patients ($n = 293$); patients with a mutation are indicated by a black rectangle. The top bar plot depicts TMB in each patient. The final rows represent the mutation status of the pathway with or without *TP53*. Percentages on the left of the plot indicate prevalence. Genes with significant single-gene associations with TMB are marked by an asterisk. Mutations in cell-cycle-regulator genes are associated with TMB with inclusion of *TP53* (two-tailed t-test, $P = 4.01 \times 10^{-8}$), but not without inclusion of *TP53* (two-tailed t-test, $P = 0.0652156$; Supplementary Table 4). **l**, Mutation status in the cell-cycle-regulation (CCR) pathway by response. For each patient, it is determined whether they have any mutations in genes belonging to the CCR pathway, excluding *TP53*. Excluding *TP53*, there is no association between mutation status in the CCR pathway and response (two-sided Fisher's test, $P = 0.31104$; Supplementary Table 4). Sample sizes are given in parentheses. For box plots, centre mark is median, whiskers are minimum/maximum excluding outliers, and circles are outliers.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

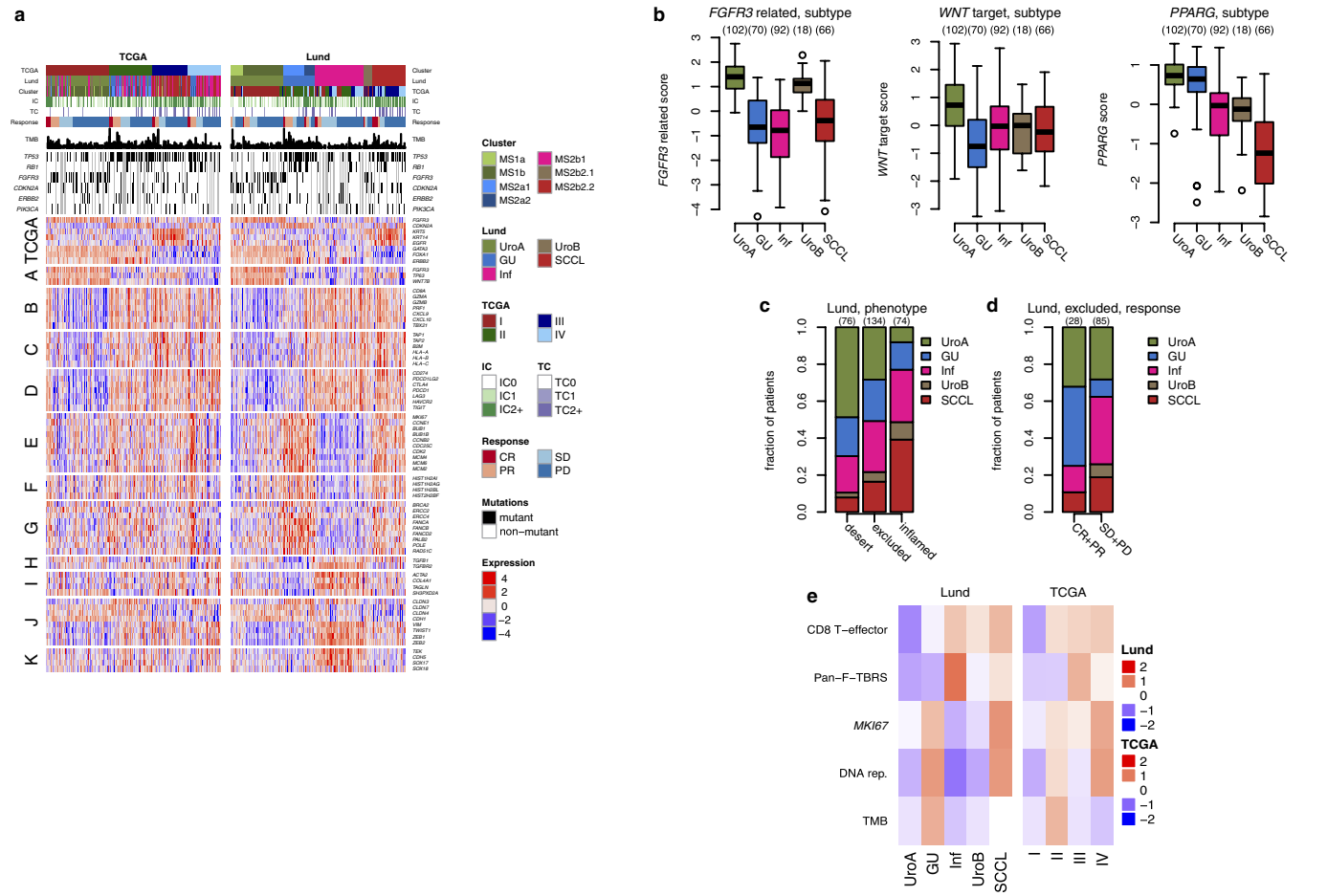


Extended Data Figure 2 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

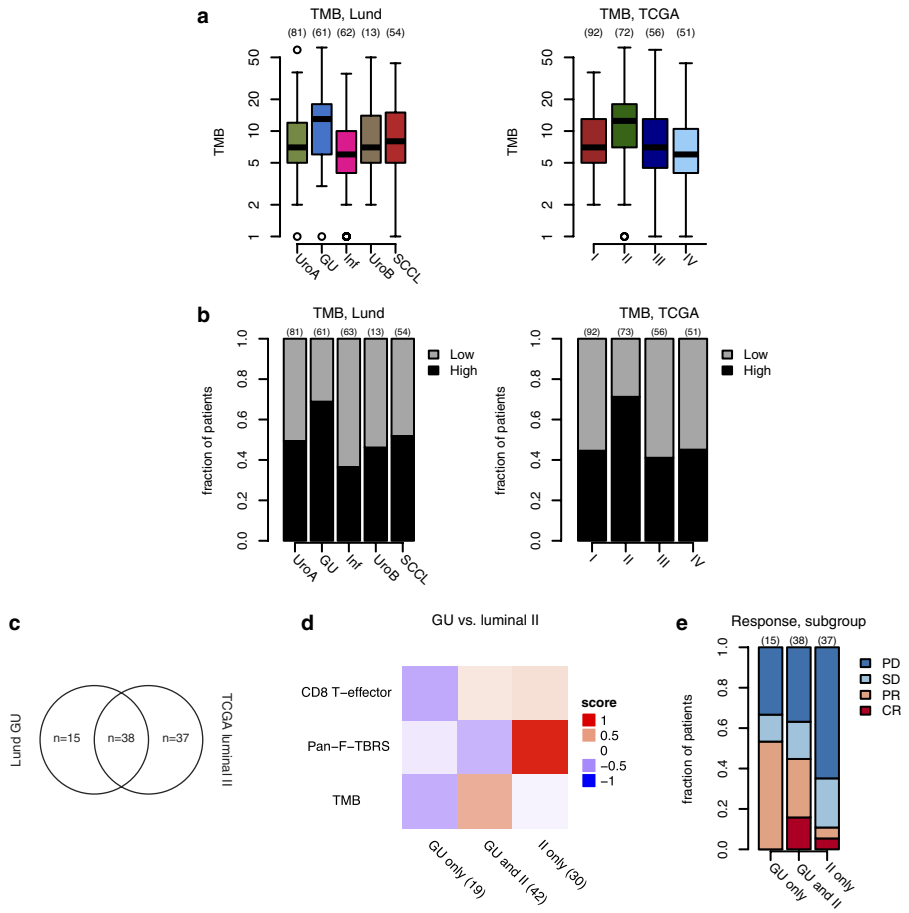
Extended Data Figure 2 | Pathways associated with response and cancer–immune phenotypes. a, KEGG pathways are significantly associated with response to atezolizumab (adjusted $P < 0.10$; comparing 68 responders to 230 non-responders). The top seven genes per set are shown; complete lists are given in Supplementary Table 6. *SGOL1* is also known as *SGO1* **b**, **c**, *IFNG* (**b**) and *IFNGR1* (**c**) gene expression (*y* axis) are significantly associated with response (two-tailed *t*-test, $P = 9.1 \times 10^{-5}$) and non-response (two-tailed *t*-test, $P = 0.00012$), respectively. **d**, **e**, *TGFBR2* gene expression (*y* axis) is significantly associated with non-response (two-tailed *t*-test, $P = 0.00019$, **d**) and, when split by quartiles, with reduced overall survival (likelihood ratio test, $P = 0.022$, **e**). In **b**–**d,** the numbers above the graphs specify sample numbers in each bin. 87 samples per quartile (**e**). **f**, Histology of tumour–immune phenotypes desert, excluded, and inflamed. **g**, Explained variance in patient response. Generalized linear models were fitted to all efficacy-evaluable, immune-phenotyped samples ($n = 204$) using binary response (complete response/partial response versus stable disease/progressive disease) as the dependent variable and scores from single input or input combinations (*x* axis) as independent variables. Percentage of explained variance of response is plotted on the *y* axis. Comparisons between different models were made using a likelihood ratio test. Horizontal bars describe likelihood ratio test results for two-biology versus three-biology model comparisons. Stacked significance symbols for two-biology models show results of likelihood ratio test comparison to the first single-biology model and separately to the second single-biology model, in the same order as the *x*-axis bar labels; for example, the T_{eff}, TMB model achieves three asterisks when compared to the T_{eff} single model, but an NS when compared to the TMB single model—due to dilution of its inflamed-specific signal in this all-sample analysis. A model that includes both DNA (TMB) and RNA markers ($CD8^+$ T_{eff} signature and F-TBRS) as well as interactions between the F-TBRS and both TMB and cancer–immune phenotype explains 50% of the variance observed in response, and it significantly improves on all single- and two-biology models. This final bar is also given on the far right in Fig. 2f. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. NS, $P > 0.1$. Exact likelihood ratio test *P* values: T_{eff},TBRS versus T_{eff}, 0.0026; T_{eff},TBRS versus TBRS, 0.0032; T_{eff},TMB versus T_{eff}, 4.9×10^{-8}; T_{eff},TMB versus TMB, 0.2; TBRS,TMB versus TBRS, 6.6×10^{-8}; TBRS,TMB versus TMB, 0.014; T_{eff},TBRS,TMB versus T_{eff},TBRS, 1.9×10^{-7}; T_{eff},TBRS,TMB versus T_{eff},TMB, 0.0028; T_{eff},TBRS,TMB versus TBRS,TMB, 0.016. T_{eff}, CD8 T-effector gene signature.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



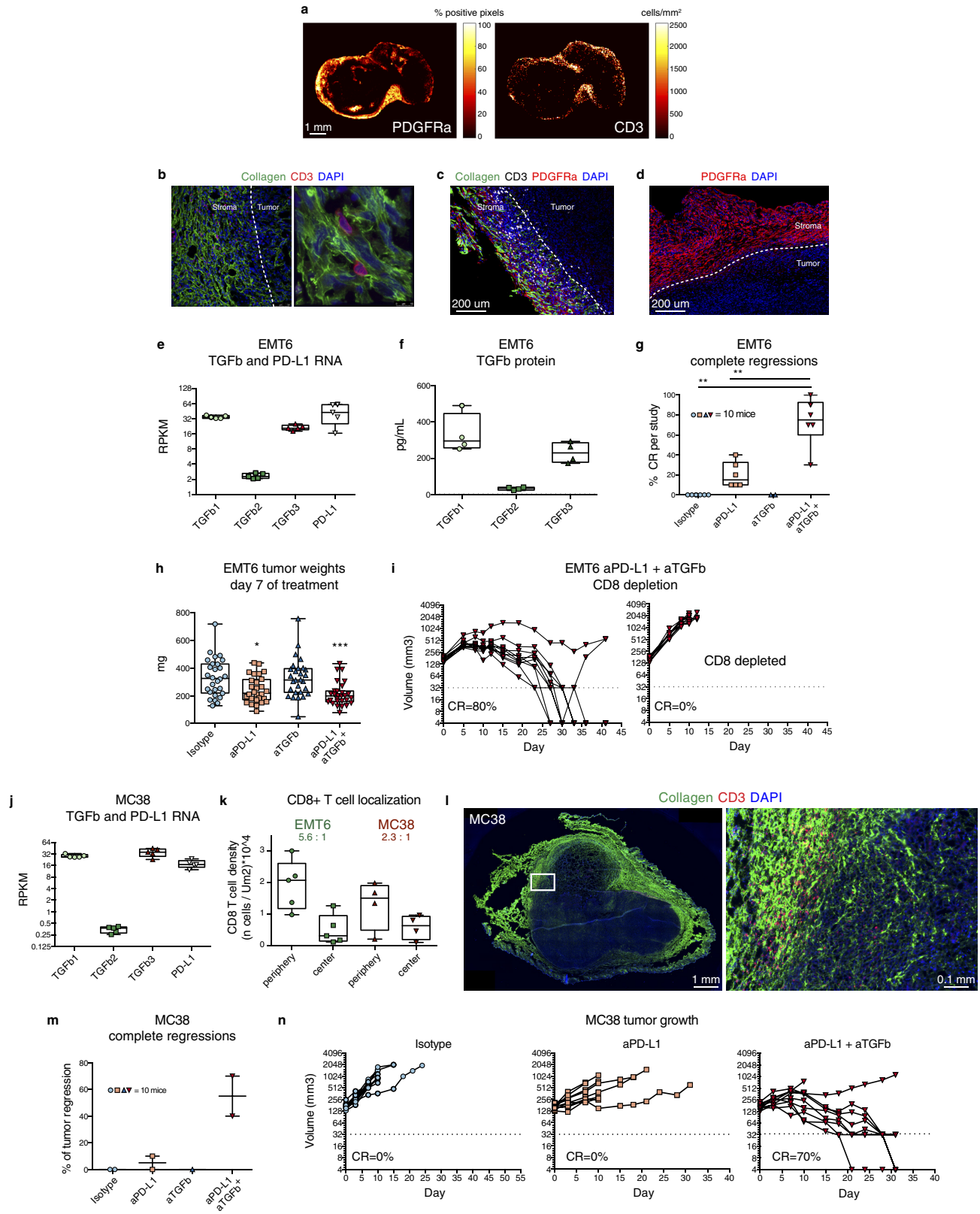
Extended Data Figure 3 | Comparison between Lund and TCGA subtyping schemes. a, Heat map representing all evaluated patients, except for patients without defined response, arranged in columns and sorted first by molecular subtype, then by response to atezolizumab. Left, patients were sorted based on a TCGA-based subtyping scheme. Right, patients were sorted by a Lund-based subtyping scheme (as in Fig. 3). Immune cell and tumour cell PD-L1 status are given. In addition, TMB and mutation status (black, mutated; grey, patients without mutation data) for a few genes of interest are shown. The rows of the heat map show expression (*Z* scores) of genes of interest, grouped into the following biologies and/or pathways: TCGA, TCGA subtyping genes; A, *FGFR3* gene signature; B, CD8 T_{eff} signature; C, antigen-processing machinery; D, immune checkpoint signature; E, *MKI67* and cell cycle genes; F, DNA replication-dependent histones; G, DNA damage-repair genes; H, TGFβ receptor and ligand; I, F-TBRS genes; J, angiogenesis signature; K, EMT markers (for details on these signatures see Supplementary Methods). **b**, *FGFR3*-related and WNT target genes[31] as well as *PPARG* are significantly differentially expressed by Lund subtype (Wald test; *P*-values: *FGFR3*-related, 2.7×10^{-43}; WNT target, 1.3×10^{-15}; *PPARG*, 1.2×10^{-53}). Gene set membership is given in Supplementary Table 8. **c**, **d**, Distribution of Lund subtypes by cancer–immune phenotypes and response status. **c**, The fraction of patients within the different Lund subtypes (*y* axis) is plotted by tumour–immune phenotype. There is a significant difference in Lund-subtype composition between cancer–immune phenotypes (χ^2 test, $P = 1.2 \times 10^{-7}$). **d**, For excluded tumours, the fraction of patients within the different Lund subtypes (*y* axis) is plotted by response status. Responders, complete and partial response (CR + PR); non-responders, stable and progressive disease (SD + PD). There is a significant difference in Lund-subtype composition between response groups (χ^2 test, $P = 0.00061$). The numbers above the graphs specify sample numbers in each bin. **e**, Assessment of *MKI67* expression and signatures of interest as well as TMB relative to molecular subtypes. Biologies of interest were scaled before medians were calculated across the Lund (left) and TCGA (right) molecular subtypes (columns). Red, high; blue, low. DNA rep., DNA replication.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 4 | Contrasting Lund and TCGA molecular subtyping. a, TMB (*y* axis) is plotted against Lund and TCGA subtypes (*x* axis). The Lund genomically unstable (two-tailed *t*-test, $P = 0.00018$) and TCGA luminal II subtypes ($P = 0.00024$) have a higher median TMB. **b**, Patients are split into TMB low (grey) and high (black), on the basis of median TMB, and the fraction of patients in these two groups is shown for the Lund and TCGA molecular subtypes. **c**–**e**, TGFβ as a probable driver of differential response in the Lund genomically unstable subtype. **c**, Three patient subgroups: Lund genomically unstable but not TCGA luminal II, both genomically unstable and luminal II, or luminal II but not genomically unstable. **d**, $CD8^+$ T_{eff}, F-TBRS and TMB by subgroup. **e**, Response differs significantly by subgroups (two-tailed Fisher's exact test, $P = 0.00062$). The numbers above the graphs or in parentheses specify sample numbers in each bin.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

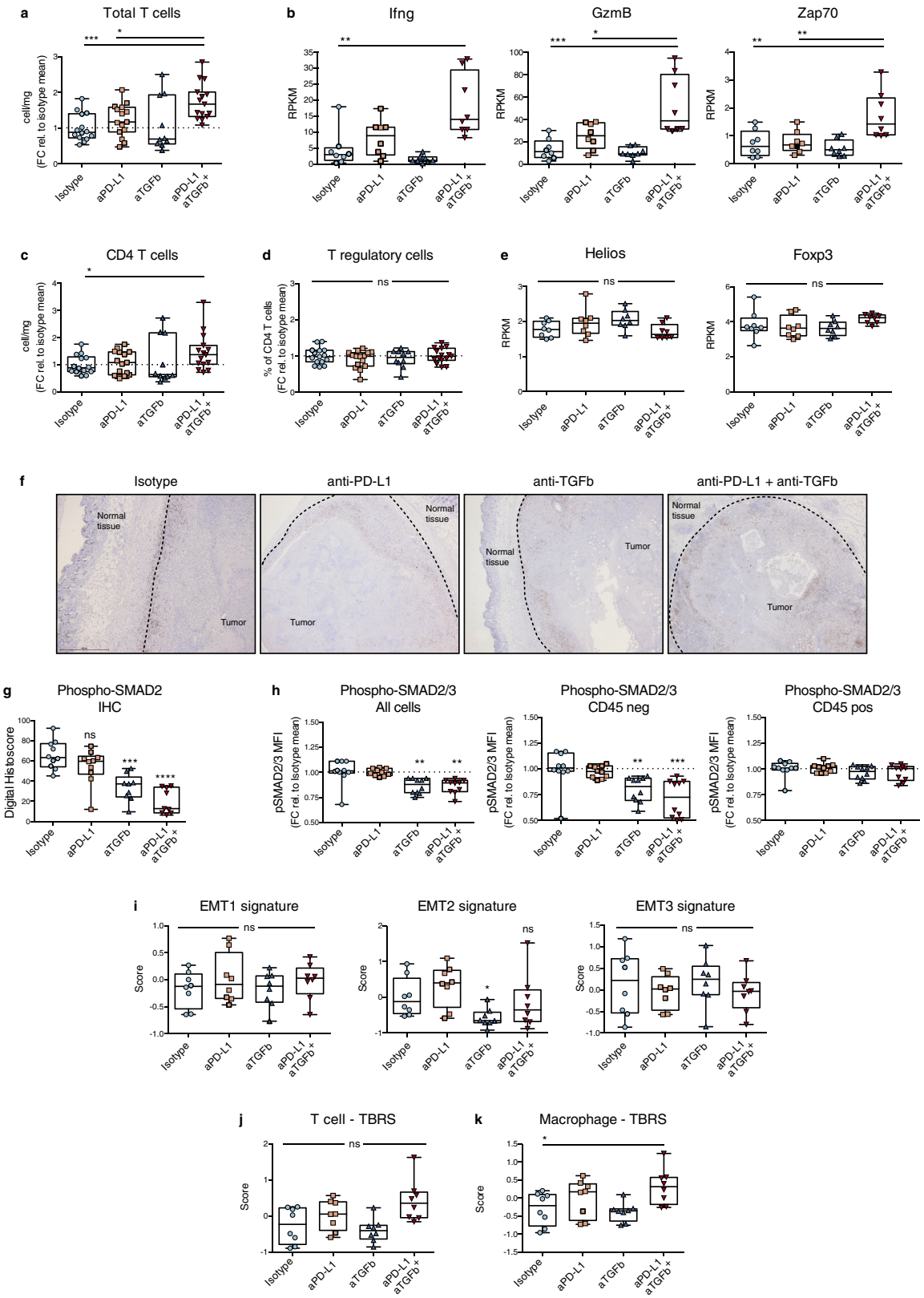


Extended Data Figure 5 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 5 | Efficacy data of anti-TGFβ and anti-PD-L1 treatment in EMT6 and MC38 immune-excluded tumour models. **a**, Fibroblast (PDGFRa, left) and T cell (CD3, right) parametric maps. Left image shows PDGFRa density (per cent positive pixels) and right shows T-cell density (cells mm^{-2}). Scale bar, 1 mm. Representative images of eight biological replicates. **b**, Collagen (green) and T cells (CD3, red) stained by immunofluorescence. Representative images of five biological replicates. **c**, Collagen (green), T cells (CD3, white) and PDGFRa (red) in EMT6 tumours stained by immunofluorescence. Scale bar, 200 μm. **d**, PDGFRa (red) in EMT6 tumours stained by immunofluorescence. Scale bar: 200 μm. Representative images of four biological replicates. **e**, Quantification of TGFβ and PD-L1 RNA in whole EMT6 tumours by RNA-seq. The tumours were inoculated orthotopically and collected when volume reached 300 mm^3 ($n = 5$ mice; data from one experiment). **f**, Quantification of TGFβ protein within whole EMT6 tumours by ELISA. Tumours were collected 14 days after inoculation, flash-frozen and lysed for protein quantification ($n = 4$ mice; data from one experiment). **g**, BALB/cJ mice were inoculated orthotopically with EMT6 tumour cells. When tumour volumes reached around 160 mm^3 approximately nine days after inoculation, mice were treated with isotype control, anti-PD-L1, anti-TGFβ, or a combination of anti-PD-L1 and anti-TGFβ. Tumours were measured two times per week for approximately eight weeks by calliper. When tumour volumes fell below 32 mm^3 (lowest limit of detection), they were considered complete response. Percentage of complete regressions across 2–6-independent studies (10 mice per group per study). **h**, Tumour weights at day seven after initiation of treatment ($n = 28$ mice per treatment group; data from three independent experiments). **i**, CD8 depletion experiment. CD8 T cells were depleted before initiation of treatment ($n = 10$ mice per group, data from one experiment). **j**, Quantification of TGFβ and PD-L1 RNA in whole MC38 tumours by RNA-seq. The tumours were inoculated subcutaneously and collected when volume reached 300 mm^3 ($n = 5$ mice; data from one experiment). **k**, Quantification of CD8 T cells in the centre and in the periphery of EMT6 and MC38 tumours from immunohistochemistry stains. Data expressed as number of cells per tissue area (periphery is defined as 400–600 μm from the tumour edge, centre is the remaining distance to the centre point). EMT6, $n = 5$ mice; MC38, $n = 4$ mice. **i**, Collagen (green) and T cells (CD3, red) in MC38 tumours stained by immunofluorescence. Scale bar, 1 mm (left); 0.1 mm (right, representative images of 5 biological replicates). **m**, C57BL/6 mice were inoculated subcutaneously with MC38 tumour cells. When tumour volumes reached around 180 mm^3 approximately eight days after inoculation, mice were treated with isotype control, anti-PD-L1, anti-TGFβ, or a combination of anti-PD-L1 and anti-TGFβ. Tumours were measured two times per week for approximately eight weeks by calliper. When tumour volumes fell below 32 mm^3 (lowest limit of detection), they were considered complete response. Percentage of complete response across two independent studies (one for anti-TGFβ alone) shown with 10 mice per treatment group for each independent study. **n**, Tumour growth curves for each individual mouse are shown. The data are representative of two independent experiments with 10 mice per treatment group. For box plots, centre mark is median, and whiskers are minimum and maximum. All statistics are two-sided Mann–Whitney *U*-tests compared to isotype group. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

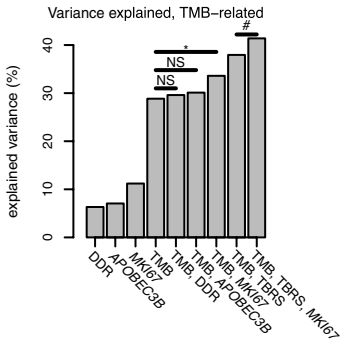


Extended Data Figure 6 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

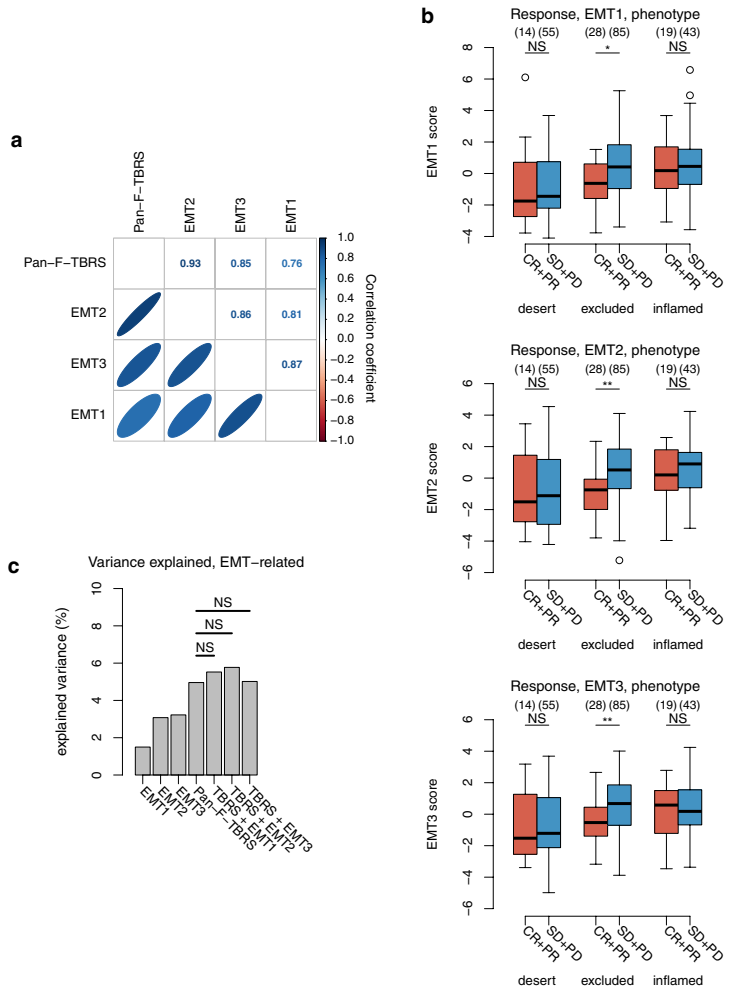
Extended Data Figure 6 | Changes in TME following anti-TGFβ + anti-PD-L1 treatment in EMT6 tumours. a, c, d, Cytofluorimetric analysis of T cells seven days after initiation of the treatment. The abundance of total T cells (**a**), total $CD4^+$ cells (**c**) and the percentage of T-regulatory cells ($CD25^+FOXP3^+$) in the $CD4^+$ population (**d**) are shown. $n = 15$ mice for all treatment groups except for anti-TGFβ alone in which $n = 10$; data combined from three independent experiments expressed as fold change relative to the isotype mean. **b**, **e**, RNA-seq analysis on whole tumours collected seven days after the initiation of treatment. Single-gene expression for *Ifng*, *Gzmb* and *Zap70* (**b**) *Ikzf2* (also known as Helios) and *Foxp3* (**e**) are shown ($n = 8$ mice per treatment group; data from one experiment). **f**, Distribution of tumour-infiltrating lymphocytes in tumours as assessed by immunohistochemistry and digital imaging seven days after the initiation of treatment as above. Representative CD3 staining (brown). Dashed line indicates tumour boundaries. ($n = 19$ for all groups except anti-PD-L1/anti-TGFβ, in which $n = 20$; three independent experiments). Scale bar, 500 μm. **g**, Quantification of pSMAD2 by immunohistochemistry at day seven after initiation of treatment. ($n = 9$ or 10 mice per treatment group; data from one experiment). **h**, Phospho-flow analysis of SMAD2/3 in tumours seven days after the initiation of treatment as above. MFI of pSMAD2/3 among total cells, $CD45^-$ or $CD45^+$ cells are shown. Data are expressed as fold change (FC) relative to the isotype MFI average. Ten mice per treatment group from two independent experiments. **i**–**k**, RNA-seq analysis on whole tumours collected seven days after the initiation of treatment. Three EMT signatures (**i**), and TGFβ-response signatures for T cells (**j**) and macrophages (**k**) are also shown. ($n = 8$ mice per treatment group; data from one experiment). All statistics in the figure use two-sided Mann–Whitney *U*-test. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. NS, not significant compared to isotype group.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 7 | Explained variance in patient response. Generalized linear models were fitted using binary response (complete or partial response versus stable or progressive disease) as the dependent variable and scores from single input or input combinations (*x* axis) as independent variables (236 samples). Percentage of explained variance of response is plotted on the *y* axis. Comparisons between different models were made using the likelihood ratio test; a significant *P* value means that the additional variable contributed some independent information to the model. The association of TMB with response is significantly stronger than that of its proxy measurements (*APOBEC3B* and *MKI67* expression or mutation in members of the DDR set). *APOBEC3B* and DDR gene set mutation provided no additional explanatory information independent of direct measurement of TMB. Combining TMB with *MKI67* expression marginally improved on TMB alone, possibly through the negative association between *MKI67* and TFGβ (Extended Data Fig. 1e, f). To test this hypothesis, we added *MKI67* to a two-pathway model based on TMB and the F-TBRS, and confirmed that *MKI67* does not add independent information to this two-pathway model. Furthermore, there was no benefit from adding *MKI67* to our full three-pathway model, shown in Fig. 2f and Extended Data Fig. 2g. #$P < 0.1$, *$P < 0.05$. Exact likelihood ratio test *P* values: TMB,DDR versus TMB, 0.38; TMB,*APOBEC3B* versus TMB, 0.26; TMB,*MKI67* versus TMB, 0.029; TMB,TBRS,*MKI67* versus TMB,TBRS, 0.064.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | Relationship between different TGFβ related gene expression signatures and response. a, Correlation between different TGFβ related gene expression signatures. Oval size and colour show correlation between signature scores and gene expression), calculated based on the complete RNA-seq dataset of 348 samples. Numbers show Pearson correlation coefficients. Gene-set membership is shown in Supplementary Table 8. See Supplementary Methods for computation of signature scores. **b**, EMT-signature expression is associated with response to atezolizumab in excluded tumours. Scores of three different EMT signatures, EMT1[32], EMT2[33] and EMT3[34], are significantly higher in non-responders (stable and progressive disease) than in responders (complete and partial response) in excluded tumours (EMT1, $P = 0.0102$; EMT2, $P = 0.0027$; EMT3, $P = 0.0063$); there is no significant difference in signature scores in desert and inflamed tumours (all $P = 1$; two-tailed *t*-test, *P* values for each signature are Bonferroni-corrected for three tests). The numbers above the graphs specify sample numbers in each bin. **c**, Explained variance in patient response. Generalized linear models were fitted using binary response (complete or partial response versus stable or progressive disease) as the dependent variable and scores from single input or input combinations (*x* axis) as independent variables (233 samples). Percentage of explained variance of response is plotted on the *y* axis. Comparisons between different models were made via likelihood ratio test; a significant *P* value means that the additional variable contributed some independent information to the model. The association of F-TBRS with response is the strongest among its correlates, that is, three different EMT signatures. None of these signatures provided additional explanatory information independent of F-TBRS.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

LETTER

doi:10.1038/nature25478

MEK drives BRAF activation through allosteric control of KSR proteins

Hugo Lavoie[1]*, Malha Sahmi[1]*, Pierre Maisonneuve[2]*, Sara A. Marullo[1], Neroshan Thevakumaran[2,3], Ting Jin[1], Igor Kurinov[4], Frank Sicheri[2,3,5] & Marc Therrien[1,6]

RAF family kinases have prominent roles in cancer[1]. Their activation is dependent on dimerization of their kinase domains, which has emerged as a hindrance for drug development[2,3]. In mammals, RAF family kinases include three catalytically competent enzymes (ARAF, BRAF and CRAF) and two pseudokinases (KSR1 and KSR2) that have been described as scaffolds owing to their apparent ability to bridge RAF isoforms and their substrate, mitogen-activated protein kinase kinase (MEK)[4]. Kinase suppressor of Ras (KSR) pseudokinases were also shown to dimerize with kinase-competent RAFs to stimulate catalysis allosterically[5]. Although GTP-bound RAS can modulate the dimerization of RAF isoforms by engaging their RAS-binding domains, KSR1 and KSR2 lack an RAS-binding domain and therefore the regulatory principles underlying their dimerization with other RAF family members remain unknown. Here we show that the selective heterodimerization of BRAF with KSR1 is specified by direct contacts between the amino-terminal regulatory regions of each protein, comprising in part a novel domain called BRS in BRAF and the coiled-coil-sterile α motif (CC-SAM) domain in KSR1. We also discovered that MEK binding to the kinase domain of KSR1 asymmetrically drives BRAF–KSR1 heterodimerization, resulting in the concomitant stimulation of BRAF catalytic activity towards free MEK molecules. These findings demonstrate that KSR–MEK complexes allosterically activate BRAF through the action of N-terminal regulatory region and kinase domain contacts and challenge the accepted role of KSR as a scaffold for MEK recruitment to RAF.

We previously developed bioluminescence resonance energy transfer (BRET) biosensors to detect interaction between RAF family kinase domains[6]. In a screen interrogating a human kinome library, we identified MEK1 and MEK2 as specific inducers of BRAF–KSR1 kinase domain dimerization (Fig. 1a, Extended Data Fig. 1a and Supplementary Table 1). We confirmed these findings by BRET and co-immunoprecipitation (co-IP) (Extended Data Fig. 1b–d). Conversely, depletion of endogenous MEK1 and MEK2 reduced the BRAF–KSR1 interaction (Extended Data Fig. 1e, f). Intriguingly, MEK promoted BRAF–KSR1 dimerization independent of its catalytic function (Fig. 1b and Extended Data Fig. 1g, h). These findings led us to assess the impact of MEK on other RAF biosensor pairs. Notably, heterodimers of BRAF–KSR1 and BRAF–KSR2 were by far the most responsive, which cannot be explained by the strict conservation of the dimer interface across RAF family members (Extended Data Fig. 1i–k)[5].

We hypothesized that MEK modulates dimer formation by interacting directly with KSR1 and/or BRAF kinase domains. Previous work showed that MEK1 binds KSR2 and BRAF in a similar 'face-to-face' manner involving their αG helices and activation segments (Extended Data Fig. 2a)[7,8]. Using yeast two-hybrid (Y2H) screening, we identified the F311S mutation in MEK1 that abrogated KSR1, BRAF, and CRAF binding (Extended Data Fig. 2b, c). In KSR1, we isolated the W831R mutation that impeded MEK1 binding (Extended Data Fig. 2d). Both mutations mapped to the contacting αG helices[7] and neither adversely affected protein fold (Extended Data Figs 2a and 3a–f). Consistent with available structures[7,8], the I666R mutation in BRAF (analogous to KSR1^{W831R}) also abolished binding to MEK1 (Extended Data Fig. 2d). We next tested the mutations for effects on BRAF–KSR1 heterodimerization. MEK1^{F311S} did not induce BRAF–KSR dimerization (Fig. 1c and Extended Data Fig. 1l), and KSR1^{W831R} could not be induced to dimerize with BRAF by MEK1 (Fig. 1d). In contrast, BRAF^{I666R} could be induced to dimerize with KSR1 by MEK1 (Fig. 1d). Similar results were obtained by co-IP using full-length proteins (Fig. 1e, compare lanes 3, 9, 11, and 13). The near-stoichiometric interaction between BRAF, KSR1, and MEK1 supported the notion that the examined interactions were strong and direct (Extended Data Fig. 3g, h). Perturbation of the side-to-side dimerization surface on KSR and BRAF kinase domains (BRAF_R509H or KSR1_R665H mutations) also abolished MEK-induced KSR1–BRAF dimerization (Fig. 1e, compare lanes 3, 5, and 7). Thus, the ability of MEK1 to promote BRAF–KSR1 association was dependent on KSR–MEK interaction and on the side-to-side dimerization surface of BRAF and KSR kinase domains[5].

We next investigated the effect of MEK1 on the catalytic output of BRAF–KSR1 dimers. MEK1 promoted the ability of KSR1 to transactivate BRAF in a kinase domain dimerization-dependent manner (Fig. 1e, phosphorylated MEK (pMEK) panel, compare lanes 3, 5 and 7, and Extended Data Fig. 4a). This held true also for KSR2 (Extended Data Fig. 4b). In contrast, when binding of MEK1 to KSR was disabled as in KSR1^{W831R}, MEK1 no longer promoted KSR1-mediated BRAF transactivation (Fig. 1e, compare lanes 3 and 9, and Extended Data Fig. 4c). Analogous findings were obtained with *Drosophila* RAF and KSR (Extended Data Fig. 4d, e). Likewise, the BRAF I666R mutation that disabled MEK1 binding also prevented MEK phosphorylation (Fig. 1e, compare lanes 3 and 11, and Extended Data Fig. 4c). Notably, MEK phosphorylation required a catalytically competent BRAF, while mutations to the active site region of KSR1 had no effect (Extended Data Fig. 4f). Finally, two KSR1 gain-of-function mutations (KSR1^{AF} and KSR1^{AFDDEE})[9] also required MEK binding for endogenous BRAF transactivation (Extended Data Fig. 4g). Thus, BRAF transactivation by KSR1 also depended on KSR–MEK interaction and on the ability of BRAF and KSR kinase domains to engage in side-to-side dimers.

We envisioned two models for how MEK1 binding to KSR1 potentiates the ability of KSR1 to promote BRAF phosphorylation of MEK1. In the first model, KSR1 functions as a scaffold to present an αG-bound MEK1 to the active site of BRAF while the kinase domains of KSR1 and BRAF engage each other as a side-to-side dimer. This would require considerable structural rearrangement because the configuration of a BRAF–KSR dimer would orient the KSR-bound MEK away from

[1]Institute for Research in Immunology and Cancer Laboratory of Intracellular Signaling Université de Montréal C.P. 6128, Succursale Centre-Ville Montréal, Québec H3C 3J7, Canada. [2]Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario M5G 1X5, Canada. [3]Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [4]NE-CAT APS, Building 436E, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, Illinois 60439, USA. [5]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [6]Département de pathologie et biologie cellulaire, Université de Montréal, Québec H3C 3J7, Canada.
*These authors contributed equally to this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

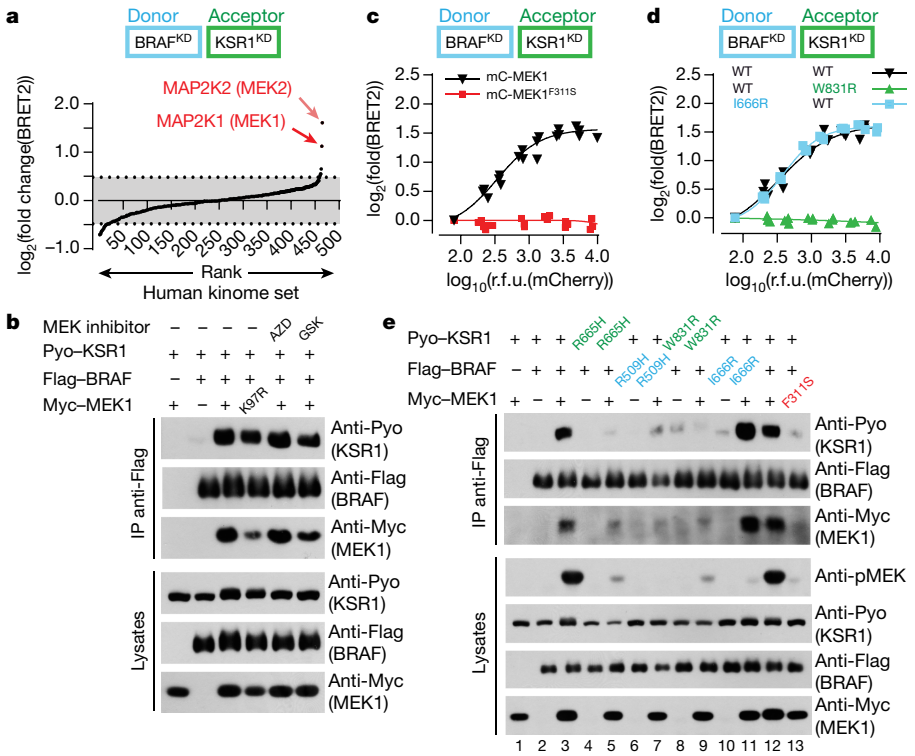


Figure 1 | MEK1 asymmetrically stimulates BRAF–KSR1 dimerization by interacting with KSR1. a, Whole-kinome screen identifies MEK1 and MEK2 (also known as MAP2K1 and MAP2K2, respectively) as specific inducers of BRAF–KSR1 dimerization. BRET biosensors comprising RlucII–BRAF^{KD}–CAAX and GFP10–KSR1^{KD}–CAAX[6] were used to screen 558 human kinase-related open reading frames (ORFs) expressed from lentiviral vectors. BRAF^{KD}, BRAK kinase domain; KSR1^{KD}, KSR1 kinase domain. **b**, Kinase-dead MEK1^{K97R} or MEK inhibitors (AZD8330 and GSK1120212) (10 μM) did not alter MEK1-induced BRAF–KSR1 dimerization in co-IPs. **c**, mCherry-tagged MEK1^{F311S} did not promote BRAF–KSR1 dimerization by BRET. r.f.u., relative fluorescence units. **d**, Heterodimerization of BRAF^{I666R} but not KSR1^{W831R} was responsive to MEK1 expression by BRET. **e**, MEK1 promotes BRAF–KSR1 dimerization and BRAF transactivation by binding KSR1 in co-IP. Experiments in **b**–**e** were repeated three times. For gel source data, see Supplementary Fig. 1.

the BRAF catalytic site. A solution to this conundrum would be the involvement in *trans* of an active BRAF dimer as previously suggested[7] and exemplified in Fig. 2a (model 1). However, as the KSR–MEK binding mode secludes the activation segment of MEK in an anti-parallel β-sheet, this too would require substantial conformational change to free the MEK phospho-acceptor sites. In a second more likely model, MEK1 binding to the KSR1 kinase domain through helix αG would support the ability of KSR1 to allosterically transactivate BRAF by side-to-side dimerization, allowing BRAF to phosphorylate a second MEK1 molecule engaged by helix αG of BRAF. This model presupposes the existence of two functionally distinct MEK molecules, namely, an 'activator' MEK that binds to KSR1 and stimulates dimerization with BRAF, and a 'substrate' MEK that is engaged and phosphorylated by activated BRAF (Fig. 2a, model 2).

To test the second model, we identified MEK1 mutants that bind KSR1^{W831R}, and thus would be functionally distinguishable from wild-type (WT) MEK1, which binds only WT KSR1 and BRAF. We reasoned that if we restored binding of MEK1 to KSR1^{W831R}, such MEK1 mutants might regain the ability to induce KSR1^{W831R} dimerization with BRAF and thus promote BRAF catalytic function. Using Y2H screening, we retrieved MEK1 variants harbouring M219V, N221Y, or M219W–A220L substitutions that restored interaction with KSR1^{W831R}, with the M219W–A220L variant to nearly WT levels (Fig. 2b, Extended Data Fig. 5a, b and Supplementary Table 2). These mutations also increased binding to endogenous KSR1, BRAF, and CRAF (Extended Data Fig. 5c). As expected, and in contrast to WT MEK1, MEK1$^{M219W–A220L}$ promoted BRAF–KSR1^{W831R} dimerization (Fig. 2c and Extended Data Fig. 5d) and induced BRAF activation (Fig. 2c). Thus, restoration of binding of MEK1 to an αG-defective KSR1 mutant was sufficient to rescue its ability to transactivate BRAF.

We took advantage of the MEK1$^{M219W–A220L}$ mutant to determine whether BRAF can phosphorylate WT MEK1 molecules not engaged directly by KSR1. We introduced a longer Myc epitope (called 8GS) at the N terminus of WT MEK1 to distinguish it by size from MEK1$^{M219W–A220L}$ (Extended Data Fig. 5e). Since WT MEK1 cannot associate with KSR1^{W831R}, it did not promote BRAF transactivation (Fig. 1e). We next co-expressed 8GS-MEK1^{WT} and MEK1$^{M219W–A220L}$ in the presence of BRAF and KSR1^{W831R}. Notably, in contrast to WT MEK1 alone, the presence of MEK1$^{M219W–A220L}$ led to the phosphorylation of 8GS-MEK1^{WT} (Fig. 2d, compare lanes 7 and 8). This required both BRAF and KSR1^{W831R} because no phosphorylation was detected in the absence of each (Fig. 2d, compare lane 8 with lanes 2, 4, and 6). Moreover, the effect was abrogated when the R665H mutation was introduced in KSR1^{W831R} (Fig. 2d, compare lanes 8 and 9). Given that 8GS-MEK1^{WT} cannot associate with KSR1^{W831R}, this result provided compelling evidence for two functional forms of MEK1: namely one molecule of MEK1 (exemplified by MEK1$^{M219W–A220L}$) that bound to KSR to enable transactivation of BRAF and a second molecule of MEK1 that functioned as a substrate of BRAF. These results supported model 2 (Fig. 2a).

We next investigated whether the ability of MEK1 to selectively induce BRAF transactivation by KSR held true for the full-length proteins. We conducted bi-directional titration experiments in the presence of MEK1 using full-length BRAF, CRAF, and KSR1. Markedly, BRAF and KSR1 co-expression led to robust pMEK levels, whereas BRAF–CRAF or CRAF–KSR1 co-expression had a marginal effect (Fig. 3a). Furthermore, MEK1 only weakly promoted full-length CRAF–KSR1 dimerization (Extended Data Fig. 6a). Despite the fact that the effect of MEK1 on dimerization was readily selective for BRAF–KSR kinase domain pairs (Extended Data Fig. 1i–k), we nonetheless questioned whether other domains participated in the context

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

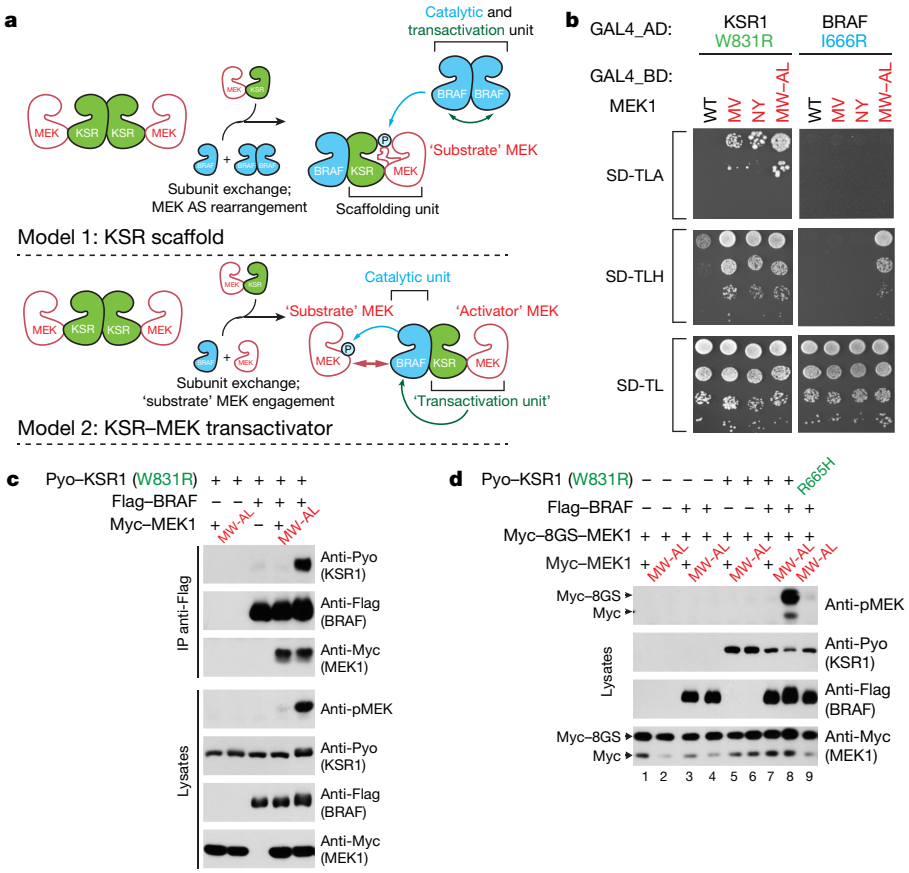


Figure 2 | KSR1-activated BRAF phosphorylates KSR1-free MEK1. **a**, Putative roles of KSR in RAF-mediated MEK phosphorylation. Model 1 was adapted from ref. 7. AS, activation segment. **b**, Y2H binding analysis of wild-type MEK1 and activation segment mutants MEK1^{M219V} (MV), MEK1^{N221Y} (NY), and MEK1$^{M219W/A220L}$ (MW–AL) to KSR1^{W831R} or BRAF^{I666R}. **c**, MEK1$^{M219W/A220L}$ (MW–AL), but not wild-type MEK1, induces KSR1^{W831R}–BRAF dimerization and transactivation as determined by co-IP and pMEK levels. **d**, MEK1$^{M219W/A220L}$ bound to KSR1^{W831R} induces BRAF activity towards MEK1 free of KSR1. Experiments in **b**–**d** were repeated at least three times. For gel source data, see Supplementary Fig. 1.

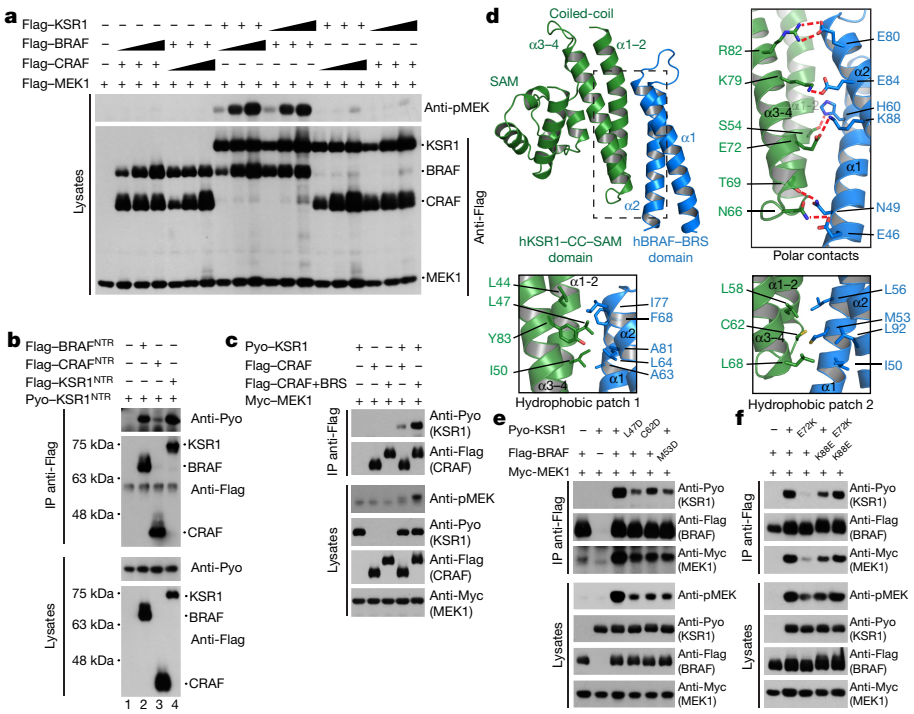


Figure 3 | The NTRs of KSR1 and BRAF selectively interact. **a**, Full-length KSR1 selectively transactivates BRAF. **b**, The KSR1 N-terminal regulatory region (NTR) associates by co-IP with itself or with the BRAF NTR, but not with the CRAF NTR. **c**, KSR1 interacts with and transactivates CRAF$^{43–648}$+BRS to a greater extent than WT CRAF. **d**, Ribbon representation of the BRS:CC-SAM complex. **e**, BRS M53D mutation and CC-SAM L47D and C62D mutations disrupt BRAF–KSR1 association and impede KSR1-induced MEK phosphorylation by BRAF. **f**, Mutational analysis of the K88^{BRS}-E72^{CC-SAM} salt bridge. Experiments in **a**–**c**, **e**, and **f** were repeated at least three times. For gel Source Data, see Supplementary Fig. 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

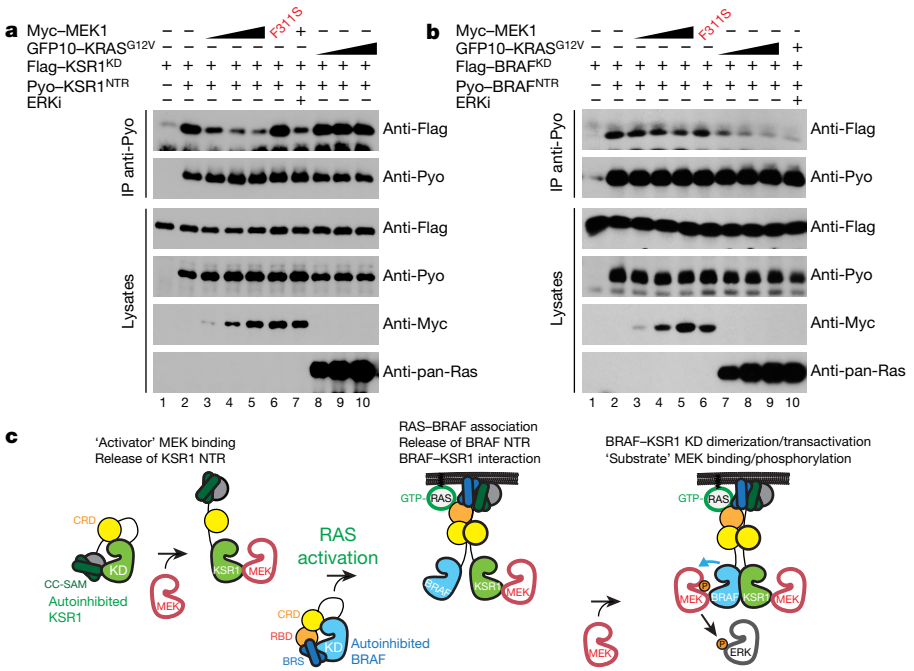


Figure 4 | MEK releases the intramolecular interaction between the NTR and kinase domain of KSR1. a, The NTR of KSR1 associates with its cognate kinase domain (KSR1^{KD}). MEK1 but not KRAS^{G12V} expression disrupts this interaction. The ERK inhibitor (ERKi) SCH772984 (10 μM) was used to exclude a potential ERK negative-feedback mechanism. **b**, The NTR of BRAF associates with its cognate kinase domain (BRAF^{KD}). Expression of KRAS^{G12V} but not MEK1 disrupts this interaction. **c**, Model of BRAF–KSR1-mediated signal transmission to MEK and ERK. Experiments in **a** and **b** were repeated at least three times. For gel source data, see Supplementary Fig. 1.

of full-length proteins as suggested previously[10]. Using co-IP, we tested the interactions between the N-terminal regulatory regions (NTRs) of BRAF, CRAF, and KSR1. Interestingly, KSR1^{NTR} interacted with BRAF^{NTR} as well as with itself, but barely with CRAF^{NTR} (Fig. 3b). The same trend was observed in BRET and Y2H analyses (Extended Data Fig. 6b, c). Thus, we concluded that the NTRs of BRAF and KSR1 also have the potential to support selective heterodimer formation. Incidentally, the NTR of KSR1 also appeared to enable KSR1 homodimerization.

We next tested NTR truncation mutants by co-IP to narrow down the domains/motifs involved. In BRAF, sequences encompassing a conserved region called the BRAF-specific (BRS) domain[3] and the cysteine-rich domain (CRD) were both required (Extended Data Figs 6d, e and 7a, b). Likewise, two areas within the KSR1 NTR encompassing the conserved CC-SAM and CRD domains were required for optimal binding (Extended Data Figs 6d, f and 7a, c). Using Y2H and BRET, we detected a specific interaction between the BRS and CC-SAM domains (Extended Data Fig. 7d, e) while the BRS did not self-associate (Extended Data Fig. 7e). Moreover, fusion of the cognate CRDs to the BRS and CC-SAM domains enhanced this interaction, whereas the isolated CRDs did not associate (Extended Data Fig. 7d). Interestingly, swapping the BRS and the CC-SAM domains between BRAF and KSR1 led to productive BRAF–KSR1 interaction and transactivation (Extended Data Fig. 7f). Conspicuously, the BRS is absent from the NTRs of ARAF and CRAF and its insertion into CRAF considerably augmented CRAF–KSR1 dimerization and CRAF transactivation (Fig. 3c and Extended Data Fig. 7a, g, h). These results identified the BRS as a novel modular element in the NTR of BRAF orthologues that confers binding selectivity towards KSR proteins via their CC-SAM domain.

Using $^{1}H–^{15}N$ heteronuclear single quantum coherence (HSQC) NMR experiments, we found that the human BRS and CC-SAM domains interacted directly (Extended Data Fig. 8a–d). By exploiting HSQC peak assignments for the mouse (m) CC-SAM domain[10], we mapped the interaction surface for the BRS to the coiled-coil motif (Extended Data Fig. 8e–g). The approximately 60-amino-acid BRS domain (Extended Data Fig. 7a, b) has no known structure. Thus, we solved its X-ray crystal structure to 2.1 Å resolution (see Extended Data Fig. 9a). The structure consisted of two anti-parallel α-helices connected by a short turn in the form of an α-hairpin (Extended Data Fig. 9b). To determine how the BRS and CC-SAM domains interact, we expressed and crystallized a BRS:CC-SAM fusion protein. The structure, solved to 1.75 Å resolution (see Extended Data Fig. 9a), revealed two BRS:CC-SAM fusion proteins in the asymmetric unit engaged in a head to tail manner (Extended Data Fig. 9c, d). This generated two BRS:CC-SAM complexes in *trans* with each mediated by extensive contact (buried surface area = 905.4 Å^2) between the coiled-coil regions in the form of a four-helix bundle (Fig. 3d and Extended Data Fig. 9c). Satisfyingly, the contact surface used by the CC-SAM domain in the crystal structure was identical to that mapped by NMR (Fig. 3d and Extended Data Fig. 8g).

The binding between the BRS and CC-SAM domains was mediated by a mixture of hydrogen bonds and salt interactions (Fig. 3d). In addition, each end of the elongated contact surface was anchored by a cluster of hydrophobic interactions (Fig. 3d). We note that binding of the BRS and CC-SAM domains correlated with major changes to the CC-SAM domain (root-mean-square deviation = 2.59 Å) and minor changes to the BRS domain (root-mean-square deviation = 0.79 Å) (Supplementary Video 1). Next, we validated the BRS:CC-SAM binding mode by mutagenesis. Point mutations targeting hydrophobic interactions in the mCC-SAM domain (L45D and C60D, corresponding to L47D and C62D in hKSR1) or in the BRS domain (M53D) impaired BRS:CC-SAM domain interaction as assessed by HSQC and by BRET (Extended Data Fig. 10a, b). The mutations also reduced the ability of full-length BRAF to associate with KSR1 and compromised MEK phosphorylation (Fig. 3e). Furthermore, single-site charge reversal mutations disrupting the E72$^{CC\text{-}SAM}$–K88^{BRS} ionic pair also abrogated interaction and activity (Fig. 3d, f and Extended Data Fig. 10c). Importantly, combination of KSR1^{E72K} with BRAF^{K88E} rescued the interaction and drove MEK phosphorylation to WT levels (Fig. 3f and Extended Data Fig. 10c). These results confirmed the relevance of the BRS:CC-SAM crystal structure for protein function *in vitro* and in cells.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

We were intrigued that full-length BRAF and KSR1 and KSR2 required MEK for heterodimerization (Fig. 1), whereas the isolated NTRs interacted independent of MEK (Extended Data Fig. 10d). Previous work showed that, in quiescent cells, the NTRs of BRAF and CRAF associate with their kinase domains to enforce a repressed state, and this was released upon RAS–GTP binding to their RAS-binding domains (RBDs)[11–13]. This self-interaction might restrain the NTRs from interacting in *trans* with other NTRs. We recently found that the NTRs and kinase domains of KSR proteins also interact[14] but the absence of an RBD would preclude a direct response of this interaction to RAS–GTP. We reasoned that the ability of KSR1 and KSR2 kinase domains to stably associate with MEK might serve as a release mechanism. Consistent with this hypothesis, we observed by co-IP that the KSR1 NTR–kinase domain interaction was progressively disrupted by MEK1 dosage (Fig. 4a). This effect was not observed with the $MEK1^{F311S}$ mutant disabled for KSR1 binding and was also independent of ERK-dependent feedback[15] as ERK inhibition had no effect (Fig. 4a). Consistent with the absence of an RBD in KSR proteins, $KRAS^{G12V}$ expression had no effect (Fig. 4a). In sharp contrast, for BRAF the NTR–kinase domain interaction was only marginally perturbed by MEK1 expression, but potently disrupted by $KRAS^{G12V}$ (Fig. 4b). These results supported the notion that, unlike BRAF, the NTR–kinase domain interaction in KSR1 was not directly responsive to RAS-GTP, but was responsive to the binding of MEK1 (Fig. 4a). Thus, binding of MEK1 to KSR1 may free its NTR for subsequent interaction with BRAF or with a second KSR1 molecule (Extended Data Fig. 10e, f).

Collectively, these findings challenge the common view that KSR proteins are scaffolds that enable MEK phosphorylation by bridging MEK and RAF and instead support a model whereby KSRs function as MEK-dependent allosteric activators of BRAF (Fig. 4c). These findings also identify the RAF family NTRs as potential targets for therapeutic intervention that may afford a more favourable route than the kinase domains to selectively modulate RAS–ERK signalling in cancers.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 May; accepted 22 December 2017.
Published online 12 February 2018.

1. Desideri, E., Cavallo, A. L. & Baccarini, M. Alike but different: RAF paralogs and their signaling outputs. *Cell* **161,** 967–970 (2015).
2. Karoulia, Z., Gavathiotis, E. & Poulikakos, P. I. New perspectives for targeting RAF kinase in human cancer. *Nat. Rev. Cancer* **17,** 676–691 (2017).
3. Lavoie, H. & Therrien, M. Regulation of RAF protein kinases in ERK signalling. *Nat. Rev. Mol. Cell Biol.* **16,** 281–298 (2015).
4. Kolch, W. Coordinating ERK/MAPK signalling through scaffolds and inhibitors. *Nat. Rev. Mol. Cell Biol.* **6,** 827–837 (2005).
5. Rajakulendran, T., Sahmi, M., Lefrançois, M., Sicheri, F. & Therrien, M. A dimerization-dependent mechanism drives RAF catalytic activation. *Nature* **461,** 542–545 (2009).
6. Lavoie, H. *et al.* Inhibitors that stabilize a closed RAF kinase domain conformation induce dimerization. *Nat. Chem. Biol.* **9,** 428–436 (2013).
7. Brennan, D. F. *et al.* A Raf-induced allosteric transition of KSR stimulates phosphorylation of MEK. *Nature* **472,** 366–369 (2011).
8. Haling, J. R. *et al.* Structure of the BRAF-MEK complex reveals a kinase activity independent role for BRAF in MAPK signaling. *Cancer Cell* **26,** 402–413 (2014).
9. Hu, J. *et al.* Mutation that blocks ATP binding creates a pseudokinase stabilizing the scaffolding function of kinase suppressor of Ras, CRAF and BRAF. *Proc Natl Acad Sci USA* **108,** 6067–6072 (2011).
10. Koveal, D. *et al.* A CC-SAM, for coiled coil-sterile α motif, domain targets the scaffold KSR-1 to specific sites in the plasma membrane. *Sci. Signal.* **5,** ra94 (2012).
11. Van Aelst, L., Barr, M., Marcus, S., Polverino, A. & Wigler, M. Complex formation between RAS and RAF and other protein kinases. *Proc. Natl Acad. Sci. USA* **90,** 6213–6217 (1993).
12. Vojtek, A. B., Hollenberg, S. M. & Cooper, J. A. Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell* **74,** 205–214 (1993).
13. Zhang, X. F. *et al.* Normal and oncogenic p21*ras* proteins bind to the amino-terminal regulatory domain of c-Raf-1. *Nature* **364,** 308–313 (1993).
14. Jin, T. *et al.* RAF inhibitors promote RAS-RAF interaction by allosterically disrupting RAF autoinhibition. *Nat. Commun.* **8,** 1211 (2017).
15. Ritt, D. A., Monson, D. M., Specht, S. I. & Morrison, D. K. Impact of feedback phosphorylation and Raf heterodimerization on normal and mutant B-Raf signaling. *Mol. Cell. Biol.* **30,** 806–819 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Volpon, M. Osborne, G. Seabrook, and the Institute for Research in Immunology and Cancer High-Throughput Screening and Proteomics facilities for experimental assistance. H.L. was supported by a Canadian Institutes for Health Research (CIHR) Banting fellowship. N.T. was supported by a CIHR scholarship. Research was supported by Impact Grants from the Canadian Cancer Society (702319 to M.T. and 704116 to F.S.) and by CIHR operating grants (MOP119443 and PJT152873 to M.T. and FDN143277 to F.S.). M.T. and F.S. hold Canada Research Chairs in Intracellular Signaling and in Structural Biology of Signaling. Synchrotron work at the Northeastern Collaborative Access Team beamlines was supported by National Institutes of Health grants (P41 GM103403 and S10 RR029205).

Author Contributions H.L., M.S., P.M., S.A.M., F.S., and M.T. designed the experiments. H.L., P.M., F.S., and M.T. wrote the manuscript. H.L. performed BRET assays with T.J. M.S. performed co-IP and transactivation assays. H.L. and S.A.M. conducted Y2H experiments. P.M., N.T., and I.K. performed structural and biophysical experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.T. (marc.therrien@umontreal.ca) or F.S. (sicheri@lunenfeld.ca).

Reviewer Information *Nature* thanks D. Barford, S. Hubbard and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment

Plasmids and RNA interference reagents. Mammalian expression plasmids were all made in the pCDNA3.1-Hygro backbone (Invitrogen). Pyo- and Myc-encoding sequences were added by PCR and the corresponding fusions were cloned between KpnI and XbaI. GFP10-, RlucII-, mCherry-, and Flag-encoding sequences were inserted at the 5′ or 3′ end of ORFs by cloning them between KpnI and XbaI sites in plasmids containing the relevant cassette. RAF family kinase domain BRET fusion constructs were generated by inserting the RlucII or GFP10 moieties at the N terminus of the kinase domain (ARAF$^{295–606}$, BRAF$^{442–766}$, CRAF$^{334–648}$, KSR1$^{602–921}$, and KSR2$^{657–950}$) and followed by a C-terminal KRAS CAAX-box as previously reported[6]. NTRs (BRAF$^{1–434}$and CRAF$^{1–326}$, KSR1$^{1–558}$) BRET constructs have the RlucII or GFP10 moieties fused at their C terminus. Y2H bait plasmids were based on pGBKT7 while prey plasmids were generated in pGADT7 (Clontech). Y2H constructs corresponding to KSR1$^{602–921}$, MEK1, BRAF$^{1–434}$, and CRAF$^{1–326}$ were cloned between NdeI and BamHI sites, while BRAF$^{442–766}$, CRAF$^{334–648}$, and KSR1$^{1–558}$ were cloned between NdeI and EcoRI sites. Plasmids used to express *Drosophila* RAF and KSR protein fusions were described previously[5]. For bacterial expression, the human BRS (BRAF$^{36–114}$) was cloned in a pGEX-2T-based expression vector with an N-terminal TVMV (TEV) protease-cleavable glutathione *S*-transferase (GST) tag (pGEX–TEV–BRS). Human and mouse KSR1 CC-SAM domains were cloned in pPROEX-HTa with an N-terminal TVMV (TEV) protease-cleavable 6× His tag. All mutations were inserted using a QuickChange II Site-Directed mutagenesis kit (Agilent). All constructs were fully verified by Sanger sequencing.

The Broad Institute kinase ORF library was obtained from Addgene in pDONR-223 entry vectors (http://www.addgene.org/human_kinases)[16]. pDONR-223 plasmids were purified with a QIAprep Miniprep Kit (Qiagen). cDNAs were then transferred in the Gateway-compatible pLX301 lentiviral vector (Addgene) by using LR Clonase enzymatic recombination according to the manufacturer's protocol (Invitrogen). pLX301 allows for expression of untagged kinase ORFs from a CMV promoter. The kinase ORF library in pLX301 was then purified with QIAprep Miniprep Kit (Qiagen) and arrayed in eight 96-well plates at a concentration of 10 ng ml^{-1}. Control lentiviruses (empty pLX301 and pLKO.1–CMV–TurboGFP) were included in each plate.

shRNA expression constructs were in the pLKO.1-puro backbone[17]. The sequences of the shRNA reagents used in this study were as follows: shControl (SHC002; 5′-CAACAAGATGAAGAGCACCAA-3′); shMEK1 (TRCN0000002329; 5′-GCTTCTATGGTGCGTTCTACA-3′); shMEK2 (TRCN0000007006; 5′-GACTATATTGTGAACGAGCCA-3′);

Cell culture, transfection, and lentiviral infections. HEK293T and HeLa cells were obtained from the Institute for Research in Immunology and Cancer high-throughput screening platform (University of Montréal, Montréal, Canada), which were originally purchased from Sigma-Aldrich (catalogue numbers 12022001-1VL and 93021013-1VL, respectively). They were routinely tested for mycoplasma contamination. Cells were maintained in DMEM supplemented with 10% FBS at 37 °C and under 5% CO_2. For BRET experiments, 1.5×10^5 cells per well were plated overnight in 12-well plates and transfected the next day with the appropriate constructs using PEI (25 μg ml^{-1})[18]. For co-IP experiments, 2×10^6 cells were seeded overnight in 100 mm tissue culture plates and transfected the next day with the appropriate constructs using PEI. Transduction of lentiviral particles was facilitated by polybrene addition (4 μg ml^{-1}). Cells were transduced at a multiplicity of infection of 5 for 24 h after which puromycin selection (2 μg ml^{-1}) was applied for 72 h.

Lentiviral production for shRNA expression was done as follows. Two million HEK293T cells were plated in T75 flasks. The next day, cells were transfected with 5 μg of pLKO.1-shRNA, 6 μg of pΔ8.9 (*gag*, *pol*) and 700 ng of pCMV-VSVG using PEI (25 μg ml^{-1})[18]. Viral supernatants were collected 72 h after transfection and filtered through 0.2 μm PES filters. Titres were then established on HeLa cells after puromycin selection (2 μg ml^{-1}) and crystal violet staining using standard procedures.

To produce kinase ORF library lentiviral particles, HEK293T cells were seeded in 96-well plates. The next day, cells were transfected using a Biomek FX instrument with 100 ng of pLX301-KinaseX, 100 ng of psPAX2 (*gag*, *pol*), and 10 ng of pMD2G using TransIT-LT1 transfection reagent (Mirus). Viral supernatants were harvested 72 h after transfection. Lentiviral production was confirmed by robotically establishing semiquantitative titres on HeLa cells after puromycin selection (2 μg ml^{-1}) followed by crystal violet staining using standard procedures.

BRET experiments. BRET titration and dose–response experiments were conducted as previously described[6]. Briefly, 48 h after transfection using appropriate BRET constructs, cells were washed once and resuspended in Hank's balanced salt solution (Wisent). Cells were next transferred to white opaque microtitre plates (Greiner). Total GFP10 levels were measured using an EnVision plate reader with a 400 nm excitation filter and a 510 nm emission filter. Total mCherry levels were detected on a FlexStation II (Molecular Devices) with excitation and emission peaks set at 580 and 635 nm, respectively. A final concentration of 2.5 μM of Coelenterazine 400a (Biotium) was then added to the plates and, following 15 min of incubation, BRET signals were acquired using a VICTOR Light plate reader (Perkin Elmer) equipped with BRET2 emission filter set (donor: 410 ± 70 nm; acceptor: 515 ± 20 nm). At least three independent biological replicates of each BRET experiment were performed. For all BRET experiments, individual technical replicates corresponding to three independent transfections ($n = 3$) from one representative experiment are presented.

BRET lentiviral ORF kinase library screening. HEK293T cells were robotically seeded into 96-well tissue culture plates (Corning) in DMEM without phenol red (Wisent) supplemented with 10% FBS and 1% penicillin/streptomycin supplemented with polybrene (4 μg ml^{-1}), followed immediately by robotically arraying individual kinase-encoding lentiviruses of the CCSB-Broad Institute kinome library (20 μl per well). After incubation for 24 h, media and virus suspension were aspirated and replaced with complete growth media containing puromycin (2 μg ml^{-1}) to select for kinase ORF-expressing cells. Forty-eight hours after puromycin selection, cells were passed 1:5 in 96-well white opaque plates (Greiner) in DMEM without phenol red (Wisent) supplemented with 10% FBS and 1% penicillin/streptomycin (Invitrogen). The next day, cells were transfected with appropriate BRET constructs using TransIT-LT1 transfection reagent (Mirus). Two days after transfection, BRET measurements were performed. Total GFP10 levels were measured using an EnVision plate reader (PerkinElmer) with a 400 nm excitation filter and a 510 nm emission filter. BRET signals and luciferase activity were acquired 15 min after addition of 2.5 μM Coelenterazine 400a (Biotium) using a Synergy NEO plate reader (BioTek) equipped with BRET2 filters. After BRET readings, cell viability was assessed by addition of WST-1 reagent (1:20 dilution) followed by 30 min incubation and subsequent quantification on an EnVision plate reader (PerkinElmer).

BRET calculations. BRET signals correspond to the light emitted by the GFP10 acceptor construct (515 ± 20 nm) upon addition of Coelenterazine 400a divided by the light emitted by the RlucII donor construct (410 ± 70 nm). Signals referred to as BRET2 in the text and figures correspond to total BRET2 signals measured from donor (RlucII fusion) plus acceptor (GFP10 fusion) expressing samples minus the background BRET2 signal emitted by donor alone expressing samples. Total intrinsic GFP10 (expressed as relative fluorescence units) and RlucII (relative luminescence units) signals were used as a proxy to ensure that similar protein levels were compared between analogous probes. In titration experiments where the amount of GFP10 acceptor construct varied, BRET2 (*y* axis) was plotted as a function of the ratio of total GFP10 signal in relative fluorescence units ([acceptor])/total luciferase signal in relative luminescence units ([donor]) (*x* axis: [acceptor]/[donor]). BRET-based dose–response experiments were presented as BRET2 fold-change, which was calculated by dividing the BRET2 signal emitted by cells transfected with a given construct by the BRET2 signal of control cells. For mCherry fusion dose–response experiments, BRET $\log_2$-transformed fold-changes were reported as a function of $\log_{10}$-transformed mCherry relative fluorescence units.

Co-IP and western immunoblotting. Co-IP and western immunoblotting procedures were essentially conducted as follows. To prepare cell lysates, cells were washed once in cold 1× PBS and then directly lysed on plates by adding 1 ml of Triton lysis buffer (50 mM Tris at pH 7.5, 150 mM NaC1, 10% glycerol, 0.2% Triton X-100, 1 mM EDTA, 1× phosphatase inhibitor cocktail (Sigma-Aldrich), 1 mM sodium vanadate, 20 μM leupeptin, aprotinin (0.15 U ml^{-1}), 1 mM phenylmethylsulfonyl fluoride (PMSF)). Lysing cells were incubated for 20 min at 4 °C with gentle rocking, collected and spun at 14,000*g*, 4 °C for 10 min. For co-IPs, 1 μl of anti-Flag M2 (Sigma-Aldrich) or 50 μl of anti-Pyo primary antibody (hybridoma supernatant) was added to fresh cell lysates along with 12.5 μl of protein A/G agarose beads (Calbiochem) and gently rocked at 4 °C for 4 h or overnight. Immunoprecipitates were washed three times with cold lysis buffer. IPs and cell lysates were then boiled in gel loading buffer for 5 min and resolved on 7%, 10%, or 12% SDS–PAGE depending on the molecular mass of the target proteins. Gels comprising immunoprecipitation samples were transferred to PVDF membranes (GE Healthcare), whereas gels comprising cell lysates were transferred to nitrocellulose membranes (Pall Corporation). Membranes were then blocked with 2% BSA (Sigma-Aldrich) diluted in Tris-buffered saline. Blocked membranes were then probed using appropriate primary antibodies. All antibodies were diluted in Tris-buffered saline supplemented with 0.2% Tween 20. Anti-phospho-MEK (Cell Signaling Technology; catalogue number 9121), anti-MEK1/2 (Cell Signaling Technology; catalogue number 9122), anti-BRAF (Santa Cruz; catalogue number

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

sc-9002), anti-CRAF (BD-Millipore; catalogue number 610152), anti-KSR1 (Abcam; catalogue number ab68483), anti-MEK1 (BD-Millipore; catalogue number 610121), anti-MEK2 (BD-Millipore; catalogue number 610235), anti-GFP (Santa Cruz; catalogue number sc-8334), anti-Ras (Abcam; catalogue number ab108602), anti-HSP90 (New England Biolabs; catalogue number 4877), anti-CDC37 (Santa Cruz; catalogue number sc-13129) and anti-pan-14-3-3 (Cell Signaling Technology; catalogue number 8312) were used at a 1:2,000 dilution. Anti-V5 (Invitrogen; catalogue number 46-0705) was used at a 1:5,000 dilution. Anti-Flag M2 (Sigma-Aldrich; catalogue number F1804) was used at a 1:20,000 dilution. Anti-Myc (9G10) and anti-Pyo[19] supernatants from hybridomas were used at a 1:10 dilution. Anti-haemagglutinin (12CA5) supernatant from hybridoma was used at a 1:5,000 dilution. Secondary anti-mouse and anti-rabbit-HRP (Jackson Immunoresearch Laboratories; catalogue numbers 115-035-146 and 111-035-144, respectively) were respectively used at 1:10,000 and 1:20,000 dilutions in Tris-buffered saline supplemented with 0.2% Tween 20. At least three independent biological replicates of each co-IP experiment were performed. One representative example is presented in the figures and gel source data are available in Supplementary Fig. 1.

Proteins loaded on SDS–PAGE for silver staining were immunoprecipitated using anti-Flag M2 affinity gel (Sigma). Proteins of interest were isolated from the gels and identified by mass spectrometry.

Desthiobiotin-ATP binding experiments. ATP-binding protein enrichment was performed using an ActivX ATP probe (desthiobiotin-ATP) following Pierce's Kinase Enrichment Kit (Thermo-Fisher Scientific) instructions with the exception that cell lysis was conducted using Triton lysis buffer (50 mM Tris at pH 7.5, 150 mM NaC1, 10% glycerol, 0.2% Triton X-100, 1 mM EDTA, 1× phosphatase inhibitor cocktail (Sigma-Aldrich), 1 mM sodium vanadate, 20 μM leupeptin, aprotinin (0.15 U ml^{-1}), 1 mM PMSF) and biotinylated proteins were captured on Streptavidin Sepharose High Performance Resin (GE Healthcare).

Y2H methods. Yeast cell growth, transformation, and DNA extraction were performed using standard procedures. Strains used for Y2H experiments were systematically PJ69-4a or PJ69-4α[20]. Yeast cells were grown at 30 °C. Synthetic dextrose (SD) medium was SD-Trp, SD-Leu, SD-Trp-Leu (SD-TL), SD-Trp-Leu-Ade (-TLA), or SD-Trp-Leu-His (SD-TLH) (0.67% Yeast Nitrogen Base, 2% glucose, indicated amino acids drop-out). Rich medium was YPD (1% yeast extract, 2% peptone, 2% dextrose). For Y2H interaction screening, bait (pGBKT7) and prey (pGADT7) plasmids were transformed in the haploid PJ69-4a and PJ69-4α, respectively (transformants were selected on SD-Trp for pGBKT7 and SD-Leu for pGADT7)[21]. Diploids carrying both plasmids were created by mating on YPD followed by replica plating on SD-TL, SD-TLH, or SD-TLA.

Identification of KSR1 and MEK1 binding-defective mutants was performed by reverse two-hybrid screening in haploid PJ69-4a cells containing either pGADT7-KSR1 or pGBKT7-MEK1 plasmids. By mapping individual mutations to previously solved crystal structures, we identified one substitution in each kinase domain that targeted surface-exposed residues on the αG helix (Extended Data Fig. 2b). Briefly, random mutations were initially introduced in $KSR1^{602–921}$ or in MEK1 cDNAs using error-prone PCR with Taq DNA polymerase in the presence of dITP[22]. Linearized pGBKT7 or pGADT7 empty plasmids were digested with NdeI and BamHI before *in vivo* homologous recombination with mutagenized PCR products in PJ69-4a containing the bait plasmid of interest (either pGADT7- $KSR1^{602–921}$ or pGBKT7-MEK1). For library generation, the PCR:plasmid molar ratio was set at 3:1. Yeast transformants were plated on SD-TL media at approximately 1,000 colony forming units per plate. Plates were incubated for 3 days at 30 °C and then replica-plated onto SD-TLH and SD-TLA media to screen for desired mutants. Mutant colonies were selected for auxotrophy on SD-TLH and/or SD-TLA media picked and streaked for single colonies on SD-TL. Confirmation of the mutant phenotype was performed by spotting assay on SD-TL, SD-TLH, and SD-TLA after growth to saturation in liquid SD-TL. Tenfold serial dilutions of yeast cells harbouring the indicated constructs were spotted on synthetic dextrose medium selective for Y2H plasmids (TL) or on media selective for Y2H plasmids and a physical interaction between encoded proteins (TLH) or (TLA). Before spotting, cells were rinsed twice in sterile water to avoid carry over of nutrients from SD-TL medium. Mutant plasmids were recovered with a modified Qiagen miniprep protocol with glass beads and then shuffled in chemocompetent bacteria. A single bacterial transformant was then picked and grown to saturation in LB medium containing the appropriate selection (kanamycin for pGBKT7 and ampicillin for pGADT7). Plasmid DNA was subsequently isolated and sequence-verified. Expression of mutant Y2H pGBKT7 fusions and pGADT7 fusions was determined by western immunoblotting of yeast extracts with anti-Myc and anti-HA antibodies, respectively.

The identification of MEK1 rescue variants that interact with $KSR1^{W831R}$ was also performed using the error-prone PCR mutagenesis approach coupled to the Y2H screening assay with the exception that the pGAD-$KSR1^{W831R}$ was used as the bait and the rescue mutants were selected for autotrophy on either of -TLH or -TLA medium. This approach repeatedly identified two independent point mutations (M219V and N221Y) situated between the RAF phospho-acceptor sites (S218 and S222) in the activation segment (Extended Data Fig. 5a, b) that enabled binding to $KSR1^{W831R}$ (Fig. 2b).

A saturation mutagenesis Y2H screen targeting residues 219–221 was then conducted to find tighter binders. Saturation mutagenesis of the human MEK1 activation segment was performed with a degenerated oligonucleotide with NNK–NNK–NNK sequence (N and K stand, respectively, for any base and for G/C) replacing the coding sequence of residues 219–221. This design allowed spanning of all amino-acid possibilities, while excluding two stop codons. It also diminished by eightfold the number of sequence combinations, preventing screen wobble base mutants different in base content but identical in amino-acid sequence. The degenerated sequence was incorporated by a two-step PCR to create a product corresponding to MEK1 cDNA flanked with pGBK homology arms. This library was screened against the pGAD–$KSR1^{W831R}$ bait for rescue mutants.

Protein expression and purification. WT and mutants of the BRS domain of human BRAF (residues 36–114) were expressed as TEV protease-cleavable GST fusions using a modified pGEX-2T vector. WT and mutants of the CC-SAM domain of human KSR1 (residues 32–178, denoted hCC-SAM) or mouse KSR1 (residues 25–170, denoted mCC-SAM) were expressed as TEV protease-cleavable 6× His tag fusions using a pPROEX vector. The human BRS domain/human CC-SAM domain chimaeras ($KSR1^{27–172}$–$BRAF^{36–110}$) connected by a $(GS)_4$, $(GS)_3$, or $(GS)_2$ linker were expressed as TEV-cleavable 6× His tag fusions using a pPROEX vector. The chimaeras allowed us to surmount the recurrent problem of crystallizing the isolated individual domains when using non-fused proteins in co-crystallization experiments. WT and mutant forms of the kinase domain of human MEK1 (residues 61–393, denoted MEK1-KD) were expressed as non-cleavable C-terminal 6× His tag fusions using a pET28a vector. Expression constructs were transformed into BL21-CodonPlus DE3-RIL bacteria (Agilent Technologies) for protein production. Bacterial expression was induced overnight at 18 °C with 1 mM IPTG and was performed in LB or TB or M9 minimal media supplemented with $[^{15}N]NH_4Cl$ (Cambridge Isotopes) or M9 minimal media supplemented with L-selenomethionine for biochemical experimentation, NMR spectroscopy, or crystallography purposes, respectively.

BRS bacterial pellets were resuspended in 10 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM PMSF, 5 mM 2-mercaptoethanol and lysed by homogenization. Lysates were clarified by centrifugation at 4 °C for 40 min at 18,000*g*. Proteins were bound to glutathione affinity resin (GE Healthcare), eluted by cleavage of the GST tag with TEV, concentrated, and then buffer exchanged by size-exclusion chromatography (SEC) using a Superdex75 24 ml column (GE Healthcare) equilibrated in running buffer for NMR experiment (20 mM phosphate buffer, 100 mM NaCl, 0.5 mM TCEP, pH 6.5) or running buffer for crystallography (20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM TCEP).

hCC-SAM and mCC-SAM bacterial pellets were resuspended in 50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM PMSF, 20 mM imidazole, 5 mM 2-mercaptoethanol and lysed by homogenization. Lysates were clarified by centrifugation at 4 °C for 40 min at 18,000*g*. Proteins were purified by a two-step chromatography procedure using nickel-affinity chromatography (GE Healthcare) in 50 mM HEPES pH 7.5, 500 mM NaCl, 5 mM 2-mercaptoethanol (with an elution gradient up to 500 mM imidazole), followed by TEV protease treatment and then SEC on a Superdex75 24 ml column (GE Healthcare) equilibrated with 20 mM phosphate buffer, 100 mM NaCl, 0.5 mM TCEP, pH 6.5 running buffer.

Bacterial pellets of BRS:CC-SAM domain chimaeric constructs were resuspended in 50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM PMSF, 1 mM TCEP, 5% glycerol and lysed by homogenization. Lysates were clarified by centrifugation at 4 °C for 40 min at 18,000*g*. Proteins were purified by the two-step procedure described for BRS GST fusion proteins with a final SEC buffer consisting of 20 mM HEPES pH 7.5, 100 mM NaCl, 1 mM TCEP.

hMEK1 WT and F311S bacterial pellets were resuspended in 100 mM HEPES pH 7, 250 mM NaCl, 1 mM PMSF, 2.5% glycerol, 1 mM TCEP and lysed by homogenization. The lysate was clarified by centrifugation at 4 °C for 40 min at 18,000*g*. Proteins were purified by a two-step chromatography procedure using nickel-affinity chromatography (GE Healthcare). Proteins were eluted from the nickel-affinity column with lysis buffer supplemented with 500 mM imidazole, followed by SEC on a Superdex75 24 ml column (GE Healthcare) equilibrated with 20 mM HEPES pH 7.5, 250 mM NaCl, 1 mM TCEP running buffer.

Following SEC, all protein fractions corresponding to greater than 95% purity were pooled, concentrated, and flash frozen in liquid nitrogen. Protein concentration was determined by ultraviolet–visible absorption spectroscopy at 280 nm wavelength using a NanoDrop spectrophotometer (Thermo-Fisher Scientific) and theoretical extinction coefficients.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ATP fluorescence polarization binding assay. Binding reactions were performed with 25 nM fluorophore-conjugated ATP (BODIPY–ATP) and the indicated protein concentrations in fluorescence polarization buffer consisting of 20 mM HEPES pH 7.5, 250 mM NaCl, 1 mM TCEP, 10 mM $MgCl_2$, 0.01% Brij-35, and 0.1 mg ml^{-1} BSA. Binding reactions were equilibrated for 30 min in 25 μl reactions in 384-well black flat-bottomed low flange plates (Corning, 3573). Fluorescence intensities were measured using a Synergy NEO plate reader (BioTek) and fluorescence polarization was calculated with Gen5 data analysis software. The instrument G-factor and detection gains were auto-scaled to a well containing tracer alone to correspond to a reading of 20 milli-polarization units. Dissociation constant (K_d) values were calculated by nonlinear regression analysis of fluorescence polarization values performed in GraphPad Prism 5.04 using a one-site total binding model.

Protein crystallography, data collection, and structural analysis. BRS domain was crystallized at 20 °C in sitting-drops by mixing 1 μl of protein (362.5 μM, 3.5 mg ml^{-1}) with 1 μl of mother liquor of 0.1 M Bis-Tris pH 6.5, 2.0 M ammonium sulfate. X-ray diffraction was measured on a flash-frozen crystal cryo-protected in mother liquor containing 20% glycerol at 100 K on station 24-ID-C, NE CAT beamline at the Advanced Photon Source using 1.5418 Å X-ray wavelength. Data reduction was performed using HKL2000 (HKL Research). The BRS domain structure was solved using CCP4 (ref. 23) and ARCIMBOLDO Lite[24]. Model building and refinement was performed using COOT[25] and PHENIX[26], respectively (see Extended Data Fig. 9a for data collection and refinement statistics).

Unlabelled and L-selenomethionine-labelled CC-SAM:$(GS)_4$:BRS fusion proteins were crystallized at 20 °C in sitting-drops by mixing 0.4 μl of protein (1 mM, 20 g l^{-1}) with 0.4 μl of mother liquor of 0.1 M Tris pH 7.5, 3.0 M sodium formate. X-ray diffraction was measured on a flash-frozen crystal cryo-protected in mother liquor containing 25% glycerol at 100 K on station 24-ID-C, NE CAT beamline at the Advanced Photon Source using 0.9791 Å X-ray wavelength. Data reduction was performed using the XDS package[27]. Phasing was performed by selenomethionine single-wavelength anomalous dispersion using HKL2MAP[28] and the SHELX pipeline[29]. The atomic model obtained by selenomethionine single-wavelength anomalous dispersion was then used as a search model for molecular replacement with a higher-resolution native data set using PHASER[30]. Phases were improved by density modification using ARP/wARP[31]. Model building and refinement were performed using COOT[25] and PHENIX[26], respectively. All models were validated using molprobity[32]. The data statistics and refinement details are reported in Extended Data Fig. 9a. Ramachandran statistics for all models were as follows. hBRS: 94.64% favoured, 5.36% allowed, 0% outliers; hCC–SAM–hBRS: 99.5% favoured, 0.5% allowed, 0% outliers. $(GS)_2$, $(GS)_3$, and $(GS)_4$ fusions adopted the same oligomeric state in solution (Extended Data Fig. 9e).

NMR experiments. NMR experiments were performed at the University Health Network High Field NMR facility. NMR–HSQC spectra were recorded at 20 °C on an 800-MHz spectrometer or a 600-MHz Bruker AVANCE III spectrometer (supported by the Canada Foundation for Innovation). The 800-MHz and 600-MHz spectrometers were equipped with 5-mm TCI and 1.7-mm TCI CryoProbes, respectively. All NMR samples contained 200 μM or 100 μM of [^{15}N]BRS (using 800 MHz or 600 MHz, respectively) or 200 μM or 150 μM [^{15}N] CC-SAM (using 800 MHz or 600 MHz, respectively) in 50 mM sodium phosphate pH 6.5, 100 mM NaCl, 0.5 mM TCEP, and 5% D_2O. Data processing was conducted using NMRviewJ and NMRpipe[33]. NMR spectra were analysed using Analysis[34]. Backbone resonance assignments for mouse CC-SAM were reported previously[10] (Biological Magnetic Resonance Data Bank accession number 17045). In all peak intensity analyses, HSQC peak heights were used. Binding affinity between BRS and CC-SAM domains was determined on the basis of chemical shift changes. (^{1}H–^{15}N) chemical shift perturbations (CSPs) were calculated as a weighted average $\Delta\delta_{av} = [(\Delta\delta_H)^2 + (\Delta\delta_N \times 0.15)^2]^{1/2}$. Affinities were obtained by fitting CSPs of the HSQC titration spectra using Analysis[34,35].

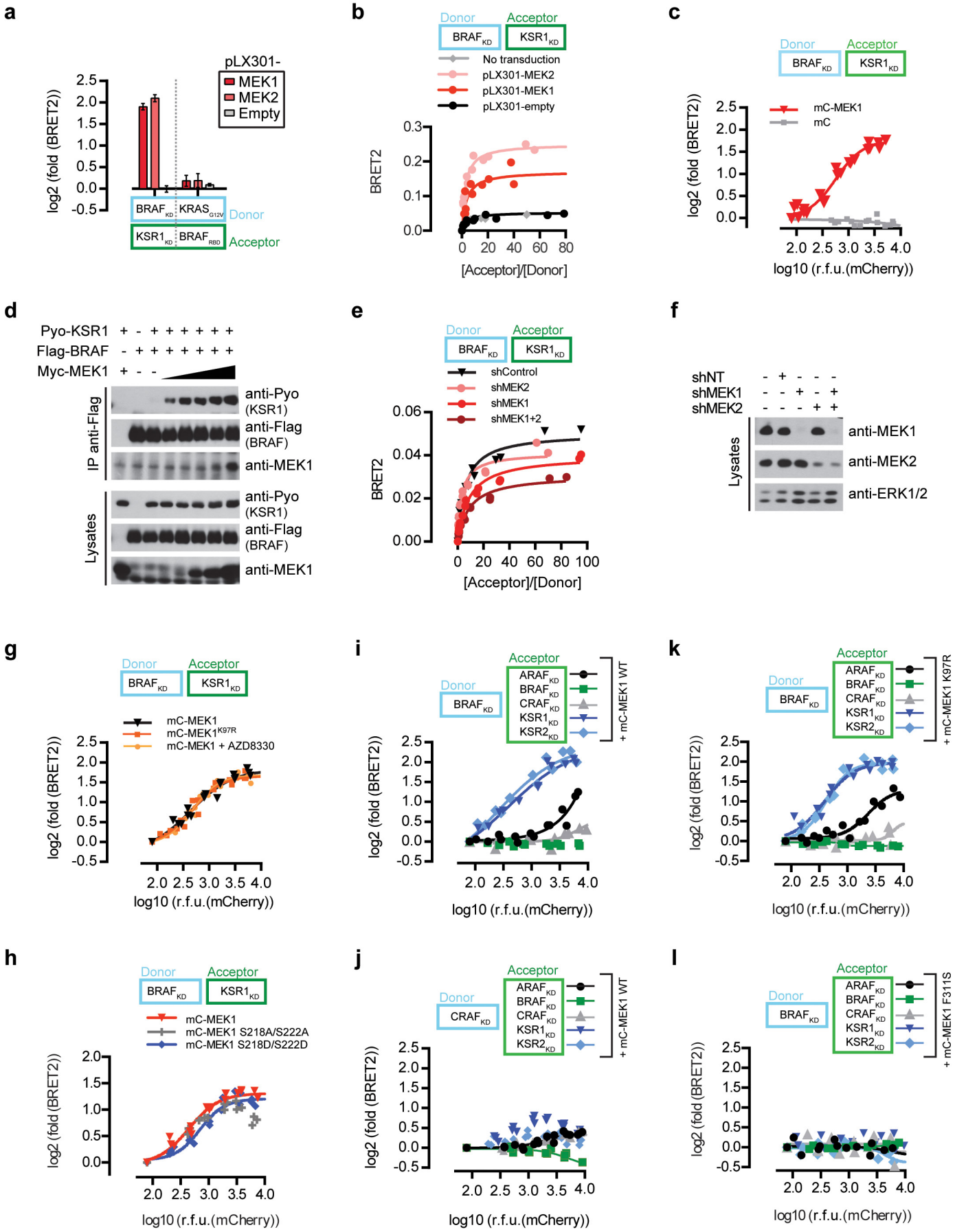
SEC followed by multi-angle light scattering analysis. CC-SAM:BRS chimaeric fusion and MEK1-KD WT and mutant proteins were analysed using a SEC followed by multi-angle light scattering (SEC–MALS) apparatus equipped with MiniDawn Treos and Optilab T-rEX detectors (Wyatt Technologies). CC-SAM:BRS chimaeric fusion samples were run at a concentration of 9 mg ml^{-1} in a buffer containing 20 mM HEPES pH 7.5, 100 mM NaCl, 1 mM TCEP. MEK1-KD samples were run at a concentration of 4 mg ml^{-1} in a buffer containing 20 mM HEPES pH 7.5, 250 mM NaCl, 1 mM TCEP. Proteins were run on a WTC030S5 (Wyatt Technologies) column at a flow rate of 0.5 ml min^{-1}. Data were processed using Astra 6.0.3.16 software (Wyatt Technologies).

Data analysis and structure rendering. BRET titration curves were analysed using Prism 5.04 (GraphPad Software) with one-site binding hyperbolic fitting of the data. BRET dose–response experiments with mCherry–MEK1 constructs were fitted using a $\log_{10}$(agonist) versus response function. Structure rendering was done using PyMol (Schrödinger)[36] and CHIMERA[37]. Coiled-coil predictions were done using the online COILS server. Sequence alignments were performed using the MUSCLE online server (http://www.drive5.com/muscle/). Alignments were edited using Jalview[38] or ALINE[39].

Data availability. All data supporting the findings of the current study are available within the Letter and its Supplementary Information files or from the corresponding authors upon reasonable request. Coordinates and structure factors for the BRS domain and the CC-SAM:BRS fusion protein have been deposited in the Protein Data Bank (PDB) under accession numbers 5VR3 and 5VYK, respectively.

16. Johannessen, C. M. *et al.* COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468,** 968–972 (2010).
17. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124,** 1283–1298 (2006).
18. Boussif, O. *et al.* A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc. Natl Acad. Sci. USA* **92,** 7297–7301 (1995).
19. Thevakumaran, N. *et al.* Crystal structure of a BRAF kinase domain monomer explains basis for allosteric regulation. *Nat. Struct. Mol. Biol.* **22,** 37–43 (2015).
20. James, P., Halladay, J. & Craig, E. A. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144,** 1425–1436 (1996).
21. James, P. Yeast two-hybrid vectors and strains. *Methods Mol. Biol*. **177,** 41–84 (2001).
22. Spee, J. H., de Vos, W. M. & Kuipers, O. P. Efficient random mutagenesis method with adjustable mutation frequency by use of PCR and dITP. *Nucleic Acids Res.* **21,** 777–778 (1993).
23. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67,** 235–242 (2011).
24. Sammito, M. *et al.* ARCIMBOLDO_LITE: single-workstation implementation and use. *Acta Crystallogr. D* **71,** 1921–1930 (2015).
25. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
26. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
27. Kabsch, W. Xds. *Acta Crystallogr. D* **66,** 125–132 (2010).
28. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.* **37,** 843–844 (2004).
29. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66,** 479–485 (2010).
30. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
31. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protocols* **3,** 1171–1179 (2008).
32. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
33. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
34. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59,** 687–696 (2005).
35. Fielding, L. NMR methods for the determination of protein-ligand dissociation constants. *Curr. Top. Med. Chem.* **3,** 39–53 (2003).
36. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8 (2015).
37. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
38. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).
39. Bond, C. S. & Schüttelkopf, A. W. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr. D* **65,** 510–512 (2009).
40. Schumann, F. H. *et al.* Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J. Biomol. NMR* **39,** 275–289 (2007).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

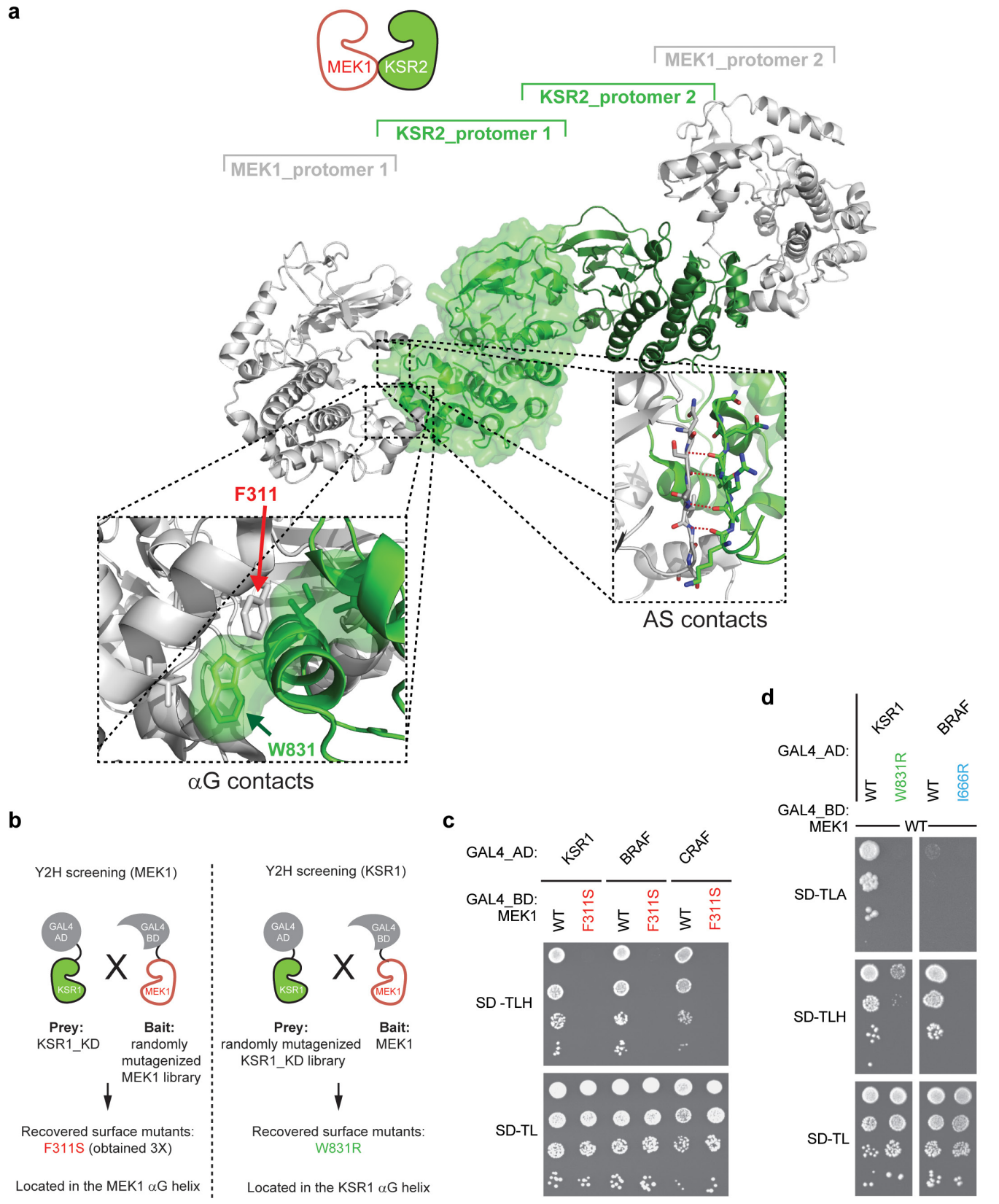


Extended Data Figure 1 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

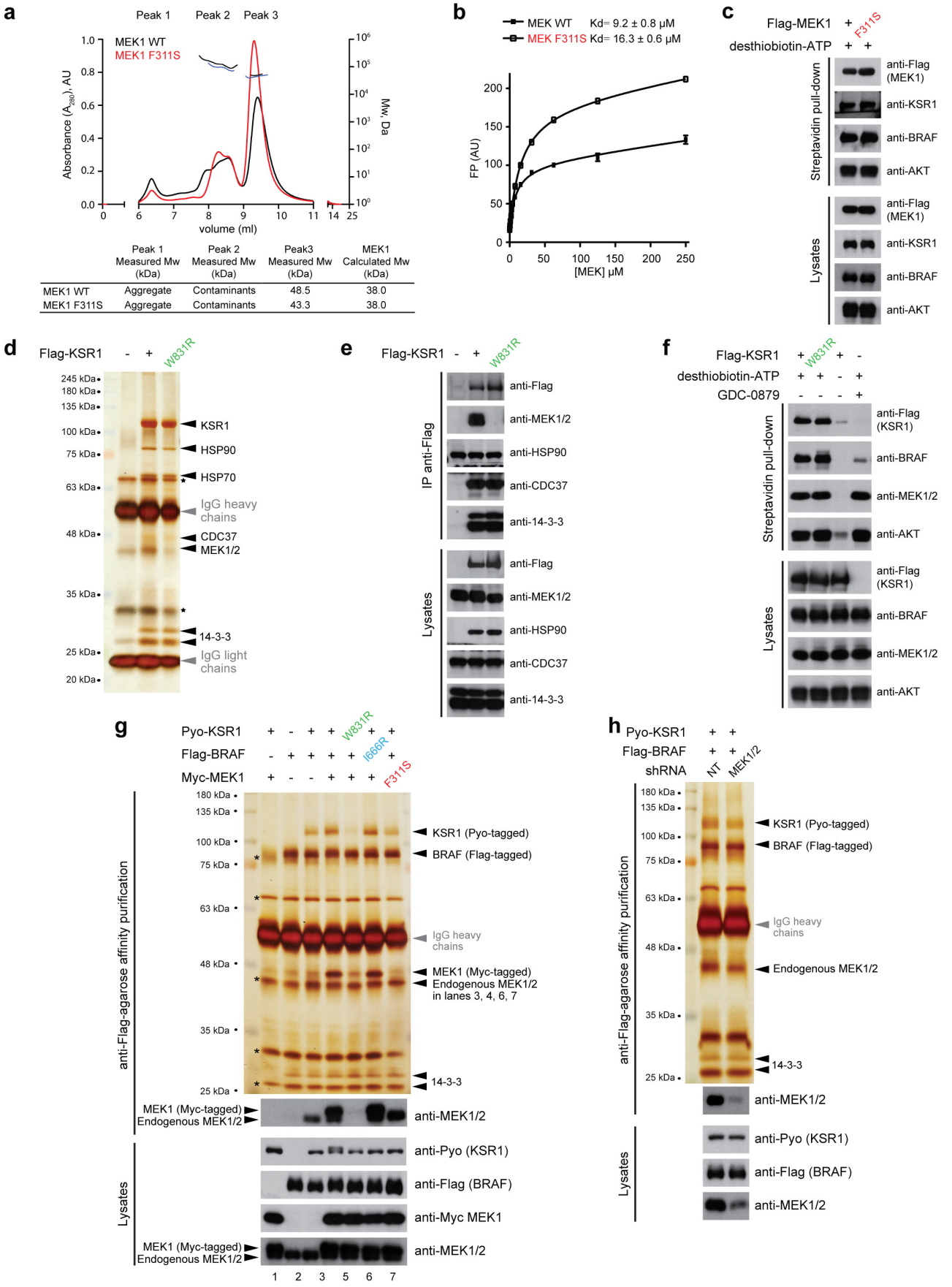
Extended Data Figure 1 | MEK levels modulate the formation of BRAF–KSR1 dimers. a, MEK1 and MEK2 expression selectively modulates BRAF–KSR1 interaction but not the $KRAS^{G12V}$–$BRAF_{RBD}$ counter screen interaction by BRET. MEK1 and MEK2 expression was driven by the lentiviral vector pLX301. Error bars correspond to mean values ± s.d. of biological triplicates. **b**, MEK1 and MEK2 modulate BRAF–KSR1 interaction by BRET. **c**, MEK1 stimulates BRAF–KSR1 interaction in a dose-dependent manner. mCherry-tagged MEK1 was titrated in cells expressing the BRAF–KSR1 BRET biosensors. **d**, MEK1 promotes full-length BRAF–KSR1 interaction by co-IP. **e**, Reducing endogenous MEK1 and/or MEK2 levels by RNA interference (shMEK1 and/or shMEK2) decreases BRAF–KSR1 interaction by BRET. **f**, Efficiency of MEK1 and MEK2 shRNA knockdowns in **d**. **g**, MEK1 stimulates BRAF–KSR1 dimerization independently of its kinase activity. mCherry-tagged MEK1 or $MEK1^{K97R}$ similarly stimulate BRAF–KSR1 interaction by BRET. Cells expressing the BRAF–KSR1 BRET biosensor and increasing amounts of WT MEK1 were treated with the MEK inhibitor AZD8330 (10 μM). **h**, MEK1 stimulates BRAF–KSR1 interaction independently of the phosphorylation status of Serine 218 and 222. mCherry-tagged MEK1, $MEK1^{S218A–S222A}$ or $MEK1^{S218D–S222D}$ similarly stimulate BRAF–KSR1 interaction by BRET. **i**, and **j**, mCherry-tagged MEK1 selectively modulates KSR-containing dimers. The impact of MEK was measured on a series of ten biosensors comprising BRAF (**h**) or CRAF (**i**) as donor probes expressed with ARAF, BRAF, CRAF, KSR1, or KSR2 as acceptor probes. **k**, mCherry-tagged kinase-dead $MEK1^{K97R}$ selectively modulates KSR-containing dimers. The impact of $MEK1^{K97R}$ was measured on a series of five biosensors comprising BRAF as the donor probe expressed with ARAF, BRAF, CRAF, KSR1, or KSR2 as acceptor probes. **l**, mCherry-tagged interaction-defective $MEK1^{F311S}$ does not modulate dimerization of five biosensors comprising BRAF as the donor probe expressed with ARAF, BRAF, CRAF, KSR1, or KSR2 as acceptor probes. In **c** and **g**–**l**, BRET $\log_2$-transformed fold-changes were reported as a function of $\log_{10}$-transformed mCherry relative fluorescence units. Experiments were repeated at least three times. For gel source data, see Supplementary Fig. 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 2 | Identification of mutations disrupting KSR–MEK and RAF–MEK interactions. a, Crystal structure of KSR2 bound to MEK1. The structure shows a heterotetrameric arrangement with two KSR2 protomers assembling as a side-to-side dimer. Each KSR2 protomer in turn interacts in a face-to-face manner with one molecule of MEK1. Zoom-in views (dashed boxes) highlight the involvement of the αG helices and the activation segments (AS) of both kinase domains in the face-to-face contact (Protein Data Bank accession number 2Y4I). **b**, Screening scheme used to identify MEK1 and KSR1 binding-defective mutants by error-prone-PCR random mutagenesis using the Y2H system. **c**, Yeast spotting assay showing that the F311S mutation in MEK1 helix αG abrogates interaction with KSR1, BRAF, and CRAF. **d**, Yeast spotting assay showing that KSR1 and BRAF helix αG mutations (W831R and I666R, respectively) disrupt their interaction with MEK1. Experiments in **c** and **d** were repeated at least three times. For gel source data, see Supplementary Fig. 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



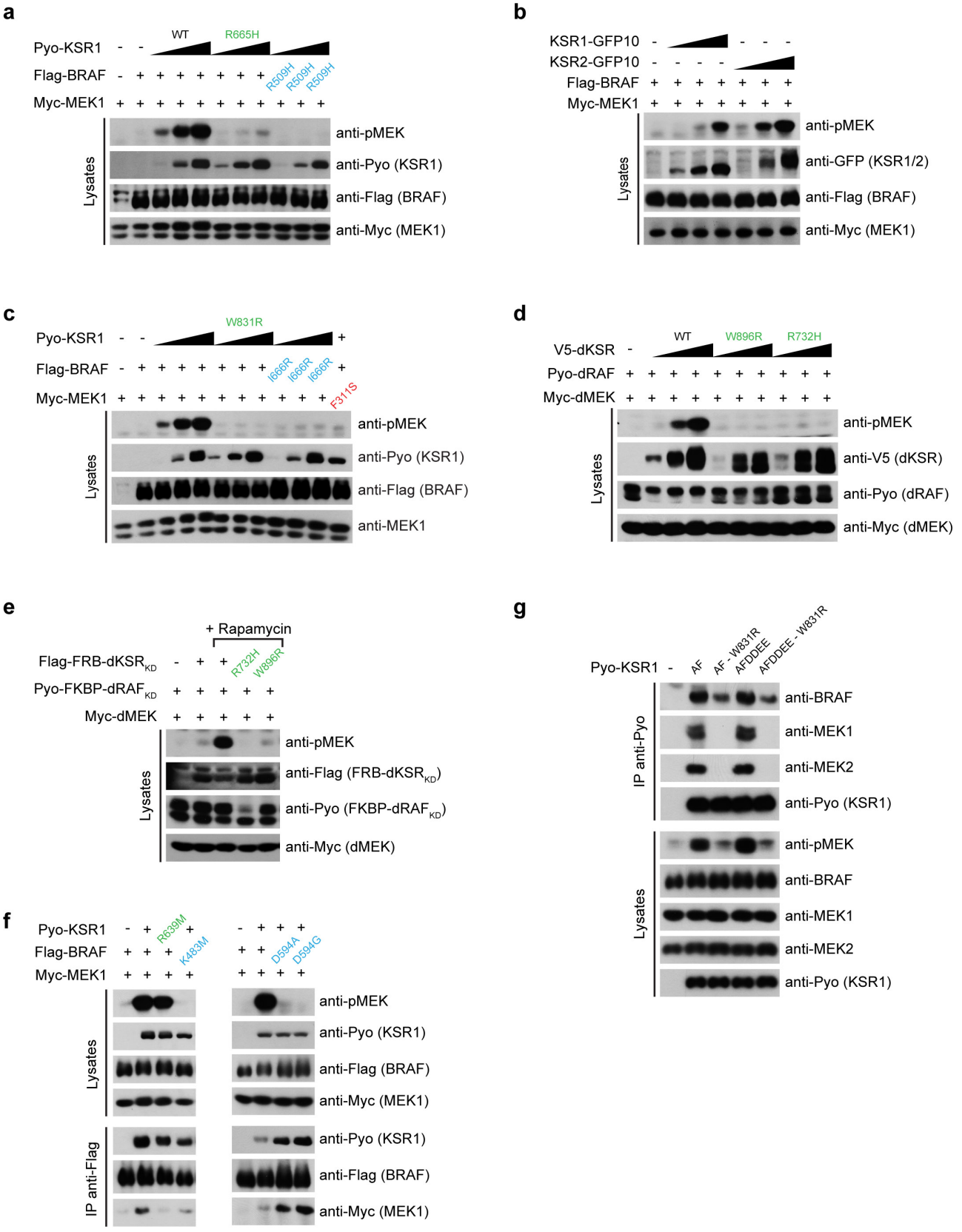
	Peak 1 Measured Mw (kDa)	Peak 2 Measured Mw (kDa)	Peak3 Measured Mw (kDa)	MEK1 Calculated Mw (kDa)
MEK1 WT	Aggregate	Contaminants	48.5	38.0
MEK1 F311S	Aggregate	Contaminants	43.3	38.0

Extended Data Figure 3 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 3 | Mutations in helix αG of MEK1 and KSR1 (F311S and W831R, respectively) do not perturb protein fold but abrogate stoichiometric assembly of the BRAF–KSR1–MEK1 complex. **a**, SEC–MALS analysis of MEK1 WT and F311S mutant. The horizontal black and blue lines correspond to SEC–MALS calculated masses for the indicated proteins. A summary table of measured and calculated molecular masses for each protein is shown below. The experiment was performed in triplicate. **b**, Fluorescence polarization (FP) binding analysis of MEK1 WT and F311S mutant to a 25 nM fluorophore-conjugated ATP tracer (BODIPY–ATP). Error bars correspond to mean values ± s.d. of technical triplicates. Indicated K_d values ± s.d. are representative of two independent experiments. **c**, MEK1 and MEK1 F311S expressed in HEK293T cells show equal ability to bind the ATP analogue desthiobiotin-ATP as assessed using a Pierce Kinase Enrichment Kit (Thermo-Fisher Scientific). **d**, **e**, WT KSR1 and the W831R mutant assemble to the same extent with the known interaction partners HSP90, CDC37, and 14-3-3, indicating that the protein fold of KSR1 is not overtly perturbed by the W831R mutation. Interaction partners were confirmed by mass spectrometry and by western immunoblotting. **f**, WT KSR1 and the W831R variant expressed in HEK293T cells show equal binding to the ATP analogue desthiobiotin-ATP as assessed using a Pierce Kinase Enrichment Kit (Thermo-Fisher Scientific). **g**, BRAF–KSR1–MEK1 complexes were purified to near-homogeneity by immunoaffinity purification on anti-Flag–agarose resin after treatment of the transfected cells with GDC-0879 (10 μM). BRAF and KSR1 form near-stoichiometric complexes in the presence of MEK1 overexpression (compare lanes 3 and 4). The KSR1 W831R mutant (lane 5) and the MEK1 F311S mutant (lane 7) but not the BRAF I666R mutant (lane 6) perturbed this near-stoichiometric interaction. Note that the BRAF–KSR1–MEK1 complex also associates at near-stoichiometric levels with endogenous 14-3-3 isoforms. In addition, in the absence of MEK1 overexpression, appreciable but sub-stoichiometric levels of endogenous MEK are detected by immunoblotting. Protein identity was determined by mass spectrometry. Asterisks indicate the position of non-specific proteins associating with the anti-Flag–agarose resin. **h**, The basal association of BRAF–KSR1 dimers (top panel, first lane) depends on endogenous MEK. Reducing MEK1 and MEK2 levels by short hairpin RNA (shRNA) interference (combination of shMEK1 and shMEK2) decreases the amount of Pyo-KSR1 co-purifying with Flag–BRAF. NT, non-target control shRNA. Experiments were repeated at least three times. For gel source data, see Supplementary Fig. 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

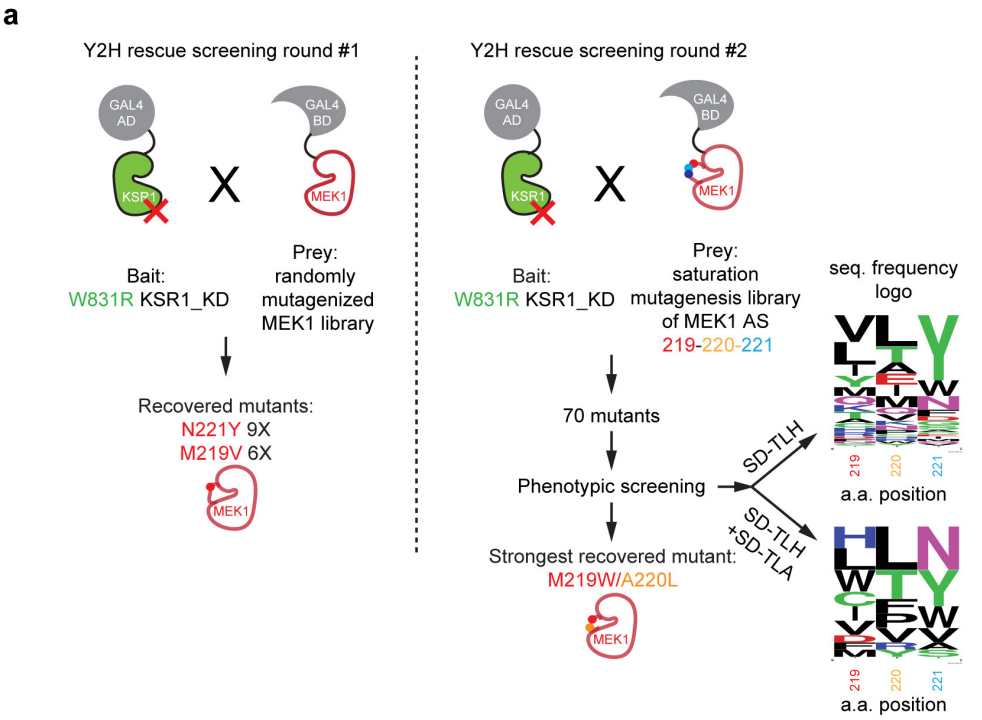


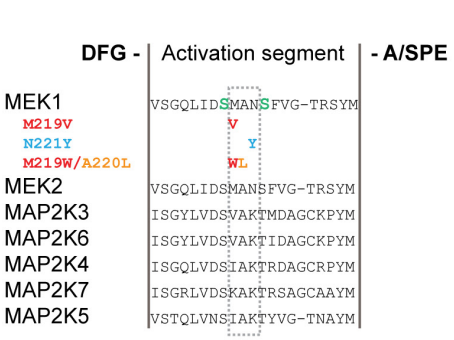
Extended Data Figure 4 | See next page for caption.

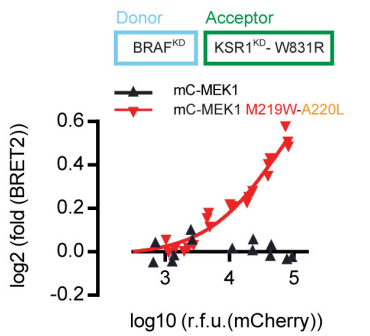
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

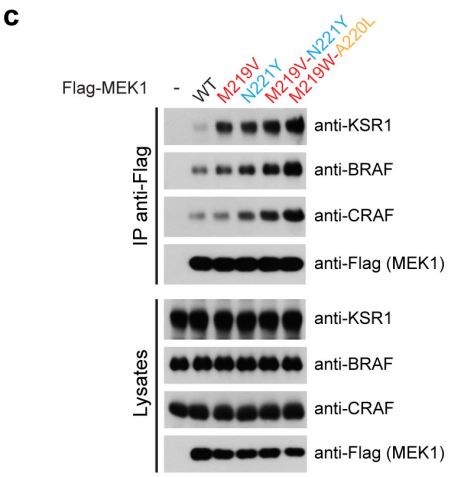
Extended Data Figure 4 | MEK1 binding to KSR1 is required for the transactivation of BRAF by side-to-side dimerization. a, BRAF activation by KSR1 depends on an intact side-to-side dimerization surface in each protomer. Co-transfection of WT BRAF and KSR1 induces a strong pMEK signal while $KSR1^{R665H}$ or $BRAF^{R509H}$ mutants do not. **b**, Human KSR1 and KSR2 can both transactivate BRAF. C-terminal GFP10 fusions of KSR1 and KSR2 were co-transfected with BRAF and the resulting pMEK signal was monitored by western immunoblotting. **c**, Disruption of KSR1–MEK1 interaction with the W831R mutation perturbs BRAF transactivation by KSR1. The BRAF I666R mutation also prevents MEK phosphorylation. Similarly, $MEK1^{F311S}$ is not phosphorylated when co-transfected with WT BRAF and KSR1. **d**, Transactivation of *Drosophila* RAF (dRAF) by *Drosophila* KSR (dKSR) requires MEK binding (W896R mutation is homologous to W831R in human KSR1) and dRAF–dKSR dimerization (R732H mutation is homologous to R665H in human KSR1). **e**, The transactivation potential of FRB–$dKSR^{KD}$ towards FKBP–$dRAF^{KD}$ was assessed by monitoring the levels of phosphorylated MEK in the presence or absence of rapamycin in S2 cells (1 μM). Blocking MEK binding to dKSR (W896R mutation) and dRAF–dKSR dimerization (R732H mutation) abrogated RAF transactivation and MEK phosphorylation as judged by western immunoblotting. **f**, MEK phosphorylation induced by co-transfection of BRAF and KSR1 depends on the catalytic activity of BRAF and not on the integrity of the active site region of KSR1. **g**, KSR1 A637F (AF) and $A637F/YLQE_{602\text{-}604}>DDEE_{602\text{-}604}$ (AFDDEE) gain-of-function mutants are dependent on binding to endogenous MEK1 or MEK2 to dimerize with endogenous BRAF and to activate the cellular pool of RAF as judged by the pMEK1 or pMEK2 signal. The W831R substitution in KSR1 was used to abrogate $KSR1^{AF}$ and $KSR1^{AFDDEE}$ mutant binding to MEK. Experiments were repeated at least three times. For gel source data, see Supplementary Fig. 1.

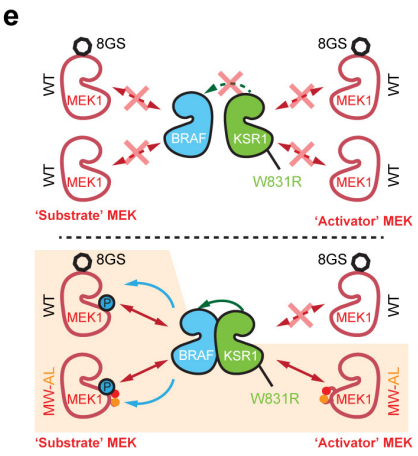
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.









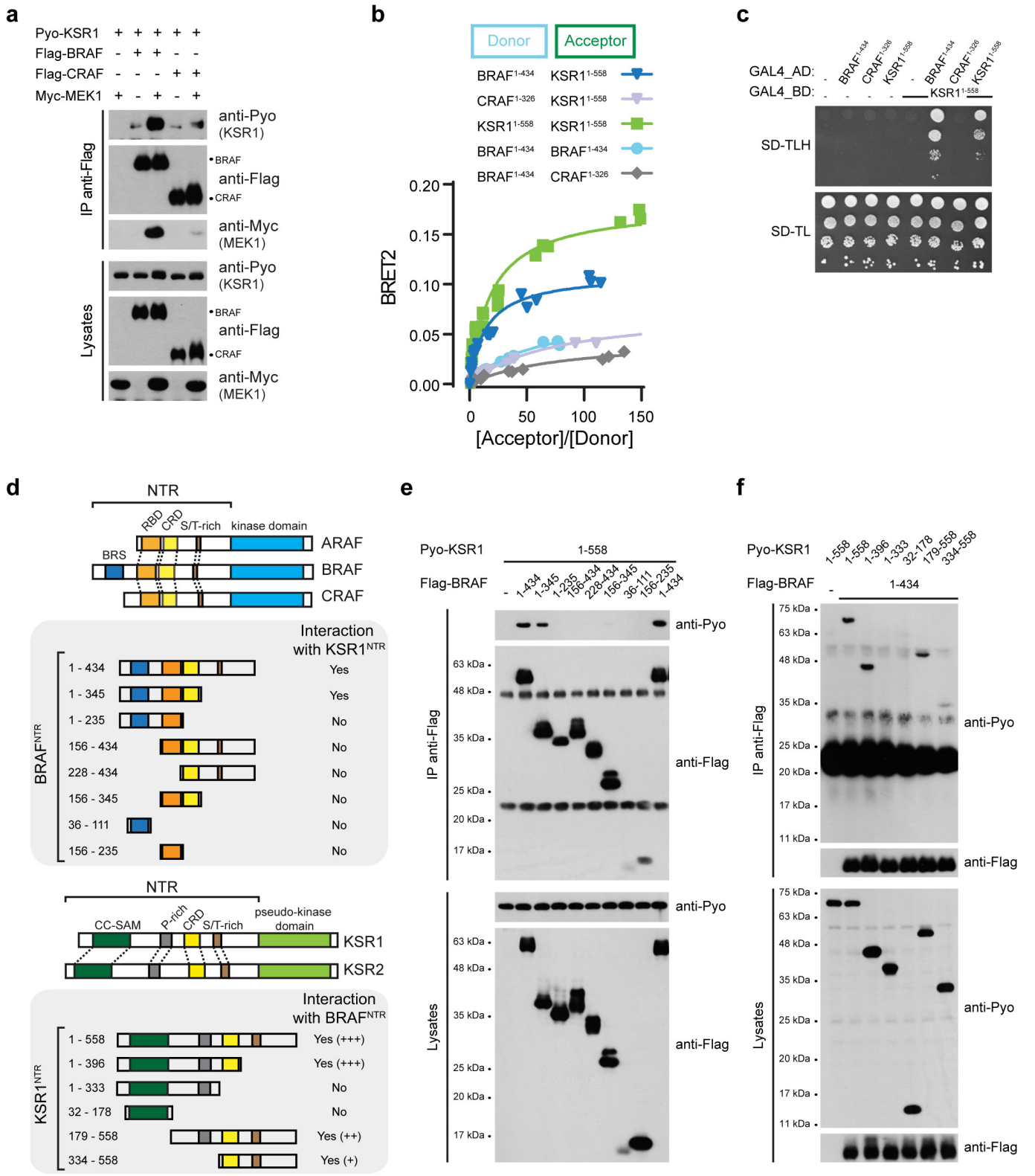


Extended Data Figure 5 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

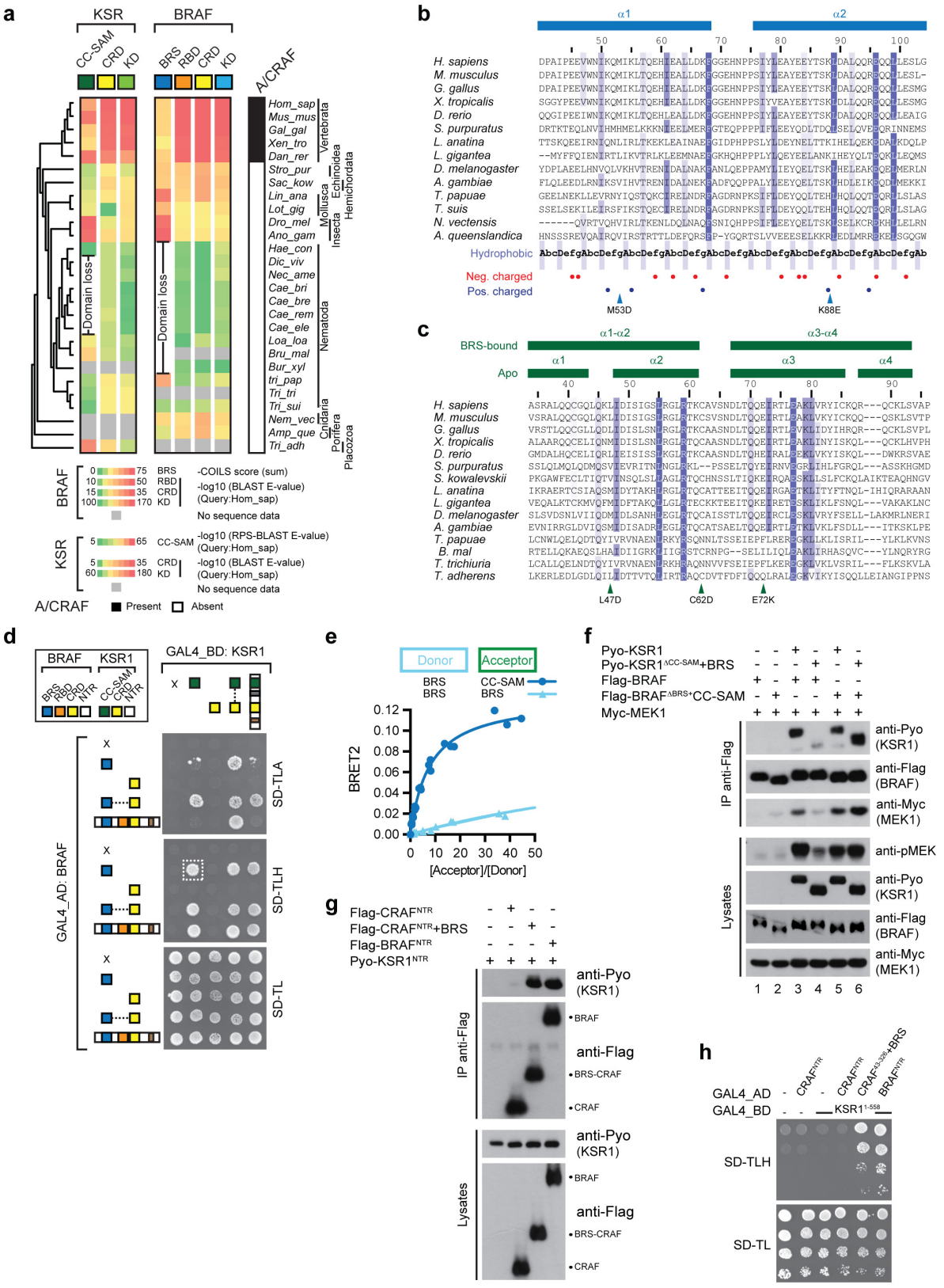
Extended Data Figure 5 | Identification of MEK1 mutants rescuing the interaction with KSR1^{W831R}. a, Screening scheme used to identify MEK1 mutants rescuing the interaction with KSR1^{W831R}. Screening was conducted in two rounds. In round 1, MEK1 was randomly mutagenized by error-prone-PCR, which identified 16 mutants corresponding exclusively to M219V and N221Y substitutions. In round 2, saturation mutagenesis was used to test all possible residue combinations in the region encoding amino acids 219–221. This identified 70 mutants that either complemented the Y2H interaction on SD-TLH or on both SD-TLH and SD-TLA. Sequence diversity of the recovered mutants is summarized by sequence logos shown on the right (see Supplementary Table 2 for details of mutant sequences). After further phenotypic screening, the fittest mutant on SD-TLA corresponded to MEK1$^{M219W–A220L}$. **b**, Position of MEK1 mutations that rescue the interaction with KSR1^{W831R}. Mutations systematically mapped to the MEK1 activation segment between Ser218 and Ser222, which are target phosphorylation sites for RAF proteins. Multiple sequence alignments of MAP2K1-7 activation segments illustrate that residues 219–221 vary within the MAP2K family, suggesting that substitutions of this sequence do not drastically affect enzyme function. The M219V substitution recovered in our screen is also found at the homologous position in MAP2K3 and MAP2K6. **c**, MEK1 activation segment mutations MEK1^{M219V}, MEK1^{N221Y}, MEK1$^{M219V/N221Y}$, and MEK1$^{M219W/A220L}$ stimulate MEK1 binding to endogenous KSR1, BRAF, and CRAF. Co-IP of Flag-tagged MEK1 variants was performed in HEK293T cells. **d**, MEK1$^{M219W–A220L}$ but not WT MEK1 stimulates KSR1^{W831R}–BRAF dimerization in BRET assays. mCherry-tagged MEK1 WT or MEK1$^{M219W–A220L}$ were titrated in cells expressing the BRAF–KSR1^{W831R} BRET biosensors. BRET $\log_2$(fold-changes) were reported as a function of $\log_{10}$-transformed mCherry relative fluorescence units. **e**, Strategy to distinguish activator and substrate MEK in BRAF transactivation assays. Eight repeats of a G_4S flexible linker (symbolized by an octagon) were added onto WT MEK1 to separate it by size from co-transfected MEK variants lacking the linker. Experimental conditions corresponding to lanes 7 and 8 of Fig. 2d, respectively, are depicted at the top and bottom of the diagram. Experiment in **e** was repeated at least three times. For gel source data, see Supplementary Fig. 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 6 | BRAF and KSR1 NTRs selectively interact. **a**, MEK expression promotes BRAF–KSR1 interaction but not CRAF–KSR1 interaction. **b**, BRET titration experiments were used to monitor the interaction between $BRAF^{NTR}$, $CRAF^{NTR}$, and $KSR1^{NTR}$. The indicated donor and acceptor pairs were tested for interaction in titration experiments. **c**, Y2H assays confirm interaction between $BRAF^{NTR}$ and $KSR1^{NTR}$, and $KSR1^{NTR}$ self-association. **d**, Domain architecture of full-length mammalian RAF and KSR isoforms and schematic of truncations used to map the molecular determinants of $BRAF^{NTR}$–$KSR1^{NTR}$ interaction. Within the NTRs are highlighted the RBDs (orange), cysteine-rich domain (CRD; yellow), serine/threonine-rich (S/T-rich; brown) region, proline-rich (P-rich; grey) region, BRAF-specific (BRS; blue) domain, and the CC-SAM (green) domain. Co-IP results derived from multiple experiments are indicated (yes/no) with representative examples shown in **e** and **f**. **e**, Co-IP of Pyo-$KSR1^{NTR}$ with the indicated fragments of Flag-tagged BRAF. **f**, Co-IP of the indicated fragments of Pyo-KSR1 with Flag-tagged $BRAF^{NTR}$. Experiments were repeated at least three times. For gel source data, see Supplementary Fig. 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

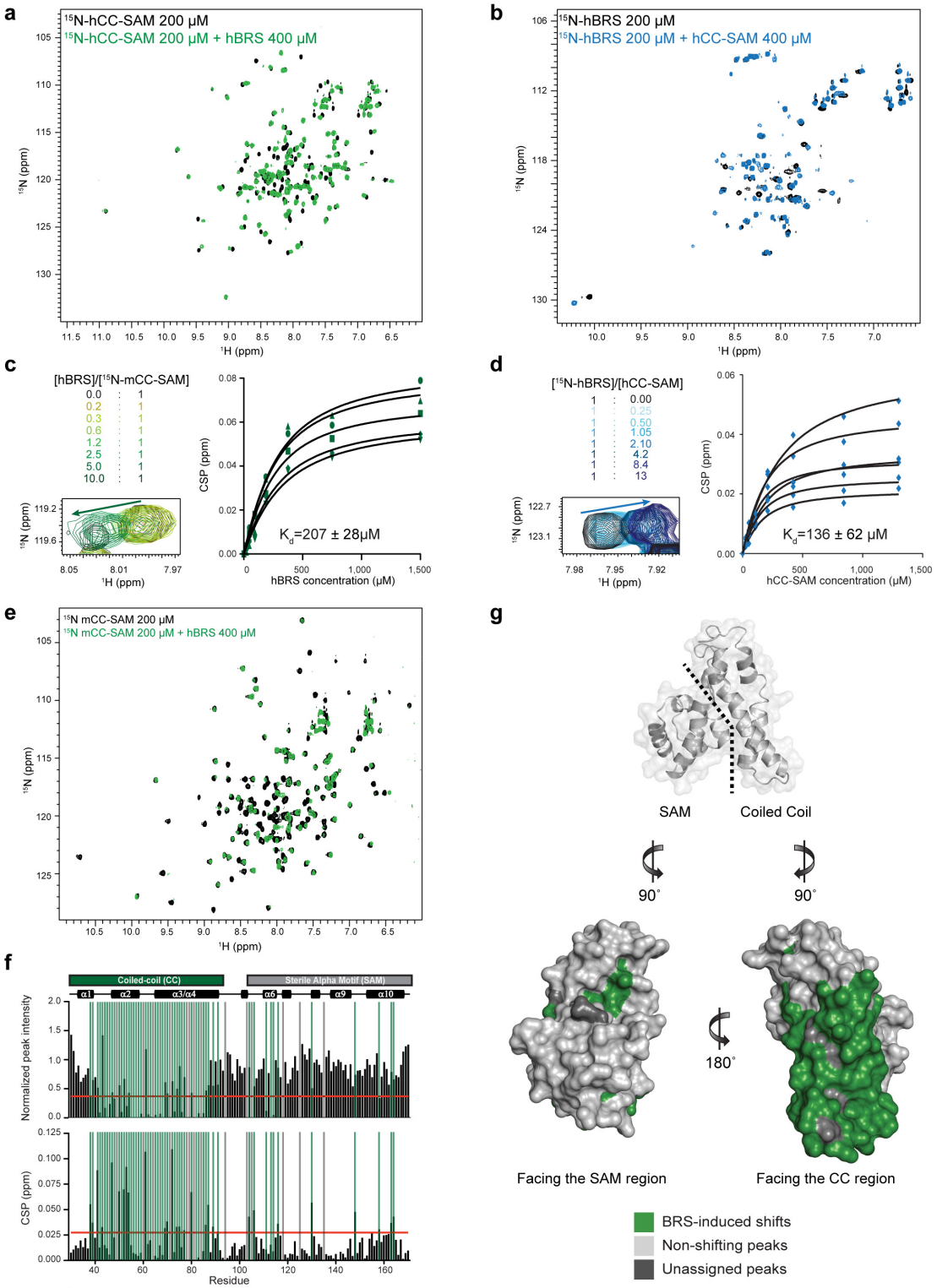


Extended Data Figure 7 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 7 | The contact surfaces between BRS and CC-SAM domains are required for the physical and functional interactions between BRAF and KSR1. a, Phylogenetic profiling of BRAF and KSR1 domains demonstrates conservation of the BRS and CC-SAM domains throughout metazoan evolution. A concomitant loss of both domains is observed in the nematode lineage (indicated by 'domain loss'). The conservation of each of BRAF and KSR1 domains with its human orthologue was colour-coded by BLAST E-value. RPS-BLAST was used to assess conservation of the CC-SAM domain. Conservation of the BRS was established by calculating the sum of the COILS score of amino acids located upstream of the RBD in each BRAF orthologue. **b**, Sequence alignment of the BRAF-specific (BRS) sequence from 14 metazoan species shows a conserved pattern of hydrophobic residues consistent with a coiled-coil region. Secondary structure elements identified in the BRS crystallographic structure are delineated above the sequence alignment. The heptad repeat (AbcDefg) typical in coiled-coil domains is shown below the alignment: positions A and D usually comprise hydrophobic amino acids while positions b, c, e, f, and g usually comprise polar or charged residues. BRS mutations used in this study are indicated by arrowheads. **c**, Sequence alignment of the CC-SAM sequence from 15 metazoan species. Secondary structure elements identified in the apo and BRS-bound CC-SAM structures are delineated above the sequence alignment. CC-SAM mutations used in this study are indicated by arrowheads below the alignment. Species in the phylogenetic tree and the sequence alignments are *Homo sapiens*, *Mus musculus*, *Gallus gallus*, *Xenopus tropicalis*, *Danio rerio*, *Strongylocentrotus purpuratus*, *Saccoglossus kowalevskii*, *Lingula anatina*, *Lottia gigantea*, *Drosophila melanogaster*, *Anopheles gambiae*, *Haemonchus contortus*, *Dictyocaulus viviparus*, *Necator americanus*, *Caenorhabditis briggsae*, *Caenorhabditis brenneri*, *Caenorhabditis remanei*, *Caenorhabditis elegans*, *Loa loa*, *Brugia malayi*, *Bursaphelenchus xylophilus*, *Trichinella papuae*, *Trichuris trichiura*, *Trichuris suis*, *Nematostella vectensis*, *Amphimedon queenslandica*, and *Trichoplax adherens*. **d**, Y2H mating matrix of BRAF and KSR1 NTR domains and chimaeric fusions (the dashed box indicates the interaction between isolated BRS and CC-SAM domains). **e**, The BRS and CC-SAM domains specifically interact as determined by BRET titration experiments. In contrast, the BRS–BRS BRET pair did not generate a saturatable BRET signal. **f**, Swapping of the BRS and CC-SAM domains supports the BRAF–KSR1 interaction and transactivation of BRAF by KSR1. Fusing the BRS domain to CRAF NTR stimulates its interaction with $KSR1^{NTR}$ by co-IP (**g**) and by Y2H (**h**). Experiments in **d**–**h** were repeated at least three times. For gel source data, see Supplementary Fig. 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 8 | NMR characterization of the physical interaction between the BRS and CC-SAM domains. a, Superposition of the [^{1}H–^{15}N]HSQC spectra of the [^{15}N]CC-SAM domain of human KSR1 (hCC-SAM) alone (black) and [^{15}N]hCC-SAM in the presence of the BRS domain of human BRAF (hBRS) (green) at a 1:2 molar ratio. Binding of BRS to CC-SAM induced both CSPs and peak broadening. **b**, Superposition of the [^{1}H–^{15}N]HSQC spectra of [^{15}N]hBRS (black) and [^{15}N]hBRS in the presence of hCC-SAM (blue) at the indicated molar concentrations. In the [^{15}N]hBRS HSQC spectra (black), numerous peaks are in intermediate (peak broadening) timescale conformational exchange, giving rise to only 50 visible peaks out of 84 expected. In the [^{15}N]hBRS:hCC-SAM HSQC spectra (blue), the binding of CC-SAM to BRS increases conformational exchange to a faster time regime, giving rise to 65 visible peaks. **c**, CSP of the mouse KSR1 [^{15}N]CC-SAM (mCC-SAM) domain as a function of titrated hBRS. Superposition of the [^{1}H–^{15}N] HSQC spectra of [^{15}N]mCC-SAM alone (black) and [^{15}N]mCC-SAM in the presence of hBRS (green) at the indicated molar concentrations (top left). The K_d value was calculated as the mean ± s.d. of five independent peak profiles (right) for which one representative CSP profile is shown (bottom left). **d**, CSP of the [^{15}N]hBRS domain as a function of titrated hCC-SAM. Superposition of the [^{1}H–^{15}N]HSQC spectra of [^{15}N]hBRS alone (black) and [^{15}N]hBRS in the presence of hBRS (blue) at the indicated molar concentrations (top left). The K_d value was calculated as the mean ± s.d. of six independent peak profiles (right) for which one representative CSP profile is shown (bottom left). **e**, Superposition of the [^{1}H–^{15}N]HSQC spectra of the [^{15}N]mCC-SAM alone (black) and in the presence of the hBRS (green) at a 1:2 molar ratio. Binding of hBRS to mCC-SAM induced both CSPs and peak broadening. **f**, Peak intensity change and CSPs versus residue number for the [^{1}H–^{15}N]HSQC spectra of [^{15}N]mCC-SAM alone versus [^{15}N]mCC-SAM with hBRS at a 1:2 molar ratio. Cut-off values indicated by the red line were calculated using the corrected s.d. method[40]. Vertical green and grey bars indicate residues identified as being part of the interacting surface and unassigned residues, respectively. **g**, Significant CSPs and peak intensity changes greater than the cut-off value determined in **f** are highlighted in green on the surface of the mCC-SAM domain (Protein Data Bank accession number 2LPE). Most CSPs and peak intensity changes map to the coiled-coil region of mCC-SAM. Non-shifting peaks are indicated in pale grey while unassigned peaks in the HSQC are coloured in dark grey.

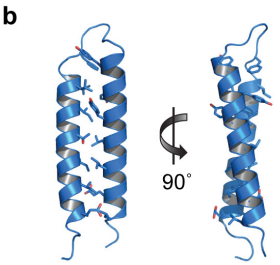
© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

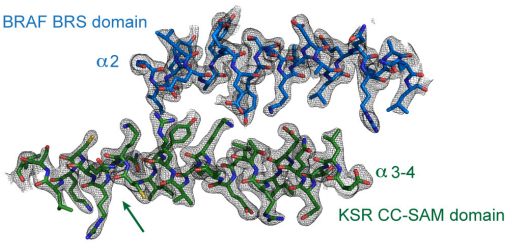
a

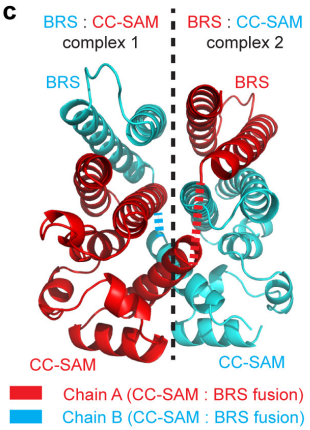
X-ray data collection and refinement statistics

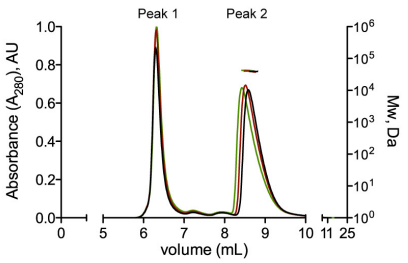
	hBRS	hCC-SAM-hBRS(SeMet)	hCC-SAM-hBRS
Data collection			
Space group	$P3_131$	$I2_12_12_1$	C121
Cell dimensions			
a, b, c (Å)	41.56, 41.56, 98.86	51.96, 54.84, 172.85	179.8, 54.34, 51.86
α, β, γ (°)	90, 90, 120	90, 90, 90	90, 106.68, 90
Resolution (Å)	50.00 - 2.10 (2.14 - 2.10)*	50.00 - 2.30 (2.34 - 2.30)	50.00 - 1.75 (1.85 - 1.75)
R_{meas}	0.056 (0.254)	0.13 (0.533)	0.04 (2.81)
$CC_{1/2}$	(0.986)	(0.891)	(0.403)
I / σI	75 (3.61)	33.9 (1.9)	14.5 (0.6)
Completeness (%)	92.8 (49.2)	96.8 (62.1)	97.9 (94.5)
Redundancy	21.9 (10.7)	8.7 (4.1)	3.9 (3.8)
Refinement			
Resolution (Å)	2.10		1.75
R_{work} / R_{free}	0.21/0.224		0.217/0.239
No. atoms			
Protein	496		3,060
Ligand/ion	5		12
Water	13		99
B-factors			
Protein	86.56		69.437
Ligand/ion	111.47		78.95
Water	72.85		61.08
R.m.s. deviations			
Bond lengths (Å)	0.009		0.006
Bond angles (°)	0.85		0.75

SeMet, selenomethionine. * : Number in parenthesis corresponds to the high resolution map.









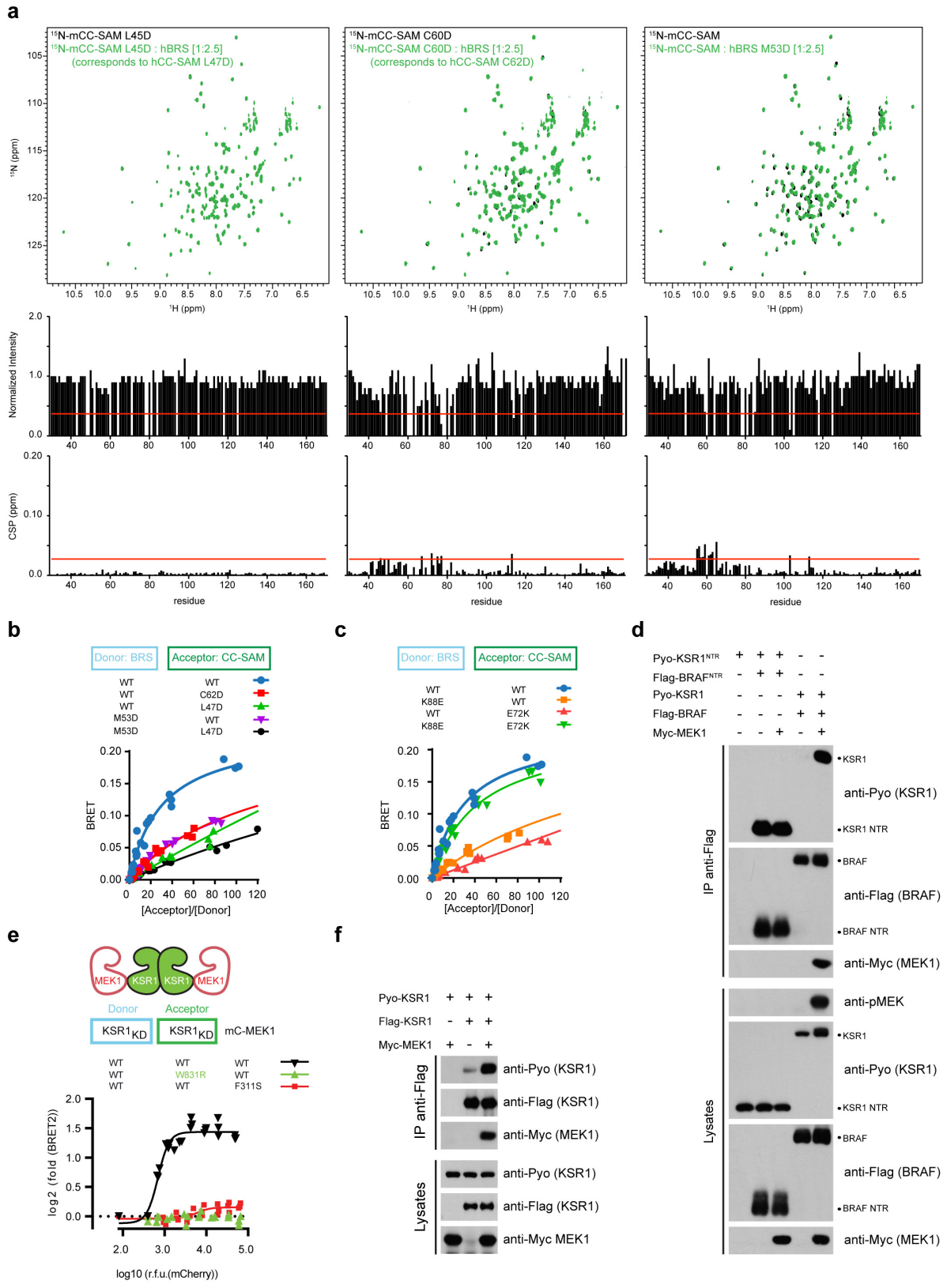
	Peak 1 Measured Mw (kDa)	Peak 2 Measured Mw (kDa)	Monomer Calculated Mw (kDa)	Dimer Calculated Mw (kDa)
CC-SAM-$(GS)_4$-BRS	Aggregate	39.5	25.6	51.1
CC-SAM-$(GS)_3$-BRS	Aggregate	39.5	25.4	50.8
CC-SAM-$(GS)_2$-BRS	Aggregate	42.0	25.3	50.5

Extended Data Figure 9 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 9 | Crystal structures of the BRS domain and the BRS:CC-SAM complex. a, X-ray data collection and refinement statistics table for the crystal structures of the hBRS domain and the hCC–SAM–hBRS complex. **b**, Ribbon representation of the crystal structure of the BRS domain of human BRAF. Residues comprising the hydrophobic core of the coiled-coil interactions are shown as sticks. **c**, Structure of the BRAF BRS and KSR1 CC-SAM domain complex. Contents of the asymmetric unit revealed two CC-SAM:BRS complexes interacting in *trans*. Disordered $(GS)_4$ linker between CC-SAM and BRS domains is represented by dashed lines. The connectivity between the BRS and the CC-SAM domains was deduced as *trans* because in this configuration, but not the *cis* configuration, the C terminus of the CC-SAM domain is sufficiently close to the N terminus of the BRS domain (15 Å) to be physically spanned by the disordered linker comprising 15 amino-acid residues. For comparison, the connection in *cis* would need to minimally span 80 Å. **d**, Representative final $2F_o - F_c$ electron density map contoured at 1.0σ for the BRS:CC-SAM complex at a region of drastic conformational rearrangement in CC-SAM relative to the isolated CC-SAM structure (indicated by the arrow). **e**, SEC–MALS analysis of the indicated crystallization fusion proteins (top). The horizontal black, red, and green lines correspond to SEC–MALS calculated masses for the indicated fusion proteins. Summary table (bottom) of measured and theoretical calculated molecular masses for each protein complex analysed. Experiments were performed in triplicate.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 10 | See next page for caption.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Figure 10 | Mutational and functional characterization of molecular interactions involving the NTRs and kinase domains of BRAF and KSR1. a, Peak intensity change and CSP versus residue number for the [^{1}H–^{15}N]HSQC spectra of the [^{15}N]mCC-SAM L45D mutant (black) and the [^{15}N]mCC-SAM L45D mutant in the presence of hBRS (green) (left), the [^{1}H,^{15}N]HSQC spectra of the [^{15}N]mCC-SAM C60D (black) versus [^{15}N]mCC-SAM C60D in the presence of hBRS (green) (middle), and the [^{1}H–^{15}N]HSQC spectra of [^{15}N]mCC-SAM (black) and [^{15}N]mCC-SAM in the presence of the hBRS M53D mutant (green) (right) at the indicated molar ratio. Cut-off values indicated by the red line were calculated using the corrected s.d. method[40]. Residue coordinates of the CC-SAM correspond to mouse KSR1; the corresponding human CC-SAM mutations are L47D and C62D. **b**, The BRS and CC-SAM domains specifically interact as determined by BRET. The CC-SAM L47D and C62D mutations reduce the BRS:CC-SAM BRET binding signal. Similarly, the BRS M53D mutation disrupts the BRS:CC-SAM BRET binding signal. **c**, Disruption of the K88–E72 salt interaction disrupts BRS:CC-SAM BRET interaction. Restoring the salt bridge with reverse polarity as in the K88E–E72K mutant pair rescues complex formation. **d**, MEK1 expression induces the formation of full-length BRAF–KSR1 heterodimers, but does not induce the formation of heterodimers between the isolated NTRs of BRAF and KSR1 as detected by co-IP. **e**, BRET analysis demonstrating that MEK expression promotes KSR1 kinase domain homodimerization. Helix αG mutations in either KSR1 (W831R) or MEK1 (F311S) abolish MEK-induced homodimerization of the KSR1 kinase domain. BRET $\log_2$-transformed fold-changes were reported as a function of $\log_{10}$-transformed mCherry relative fluorescence units. **f**, Co-IP analysis of Pyo-KSR1 with Flag–KSR1 reveals that MEK1 promotes full-length KSR1 homodimerization. Experiments in **b**–**f** were repeated at least three times. For gel source data, see Supplementary Fig. 1.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

CORRIGENDUM

doi:10.1038/nature25470

Corrigendum: mTORC1–dependent AMD1 regulation sustains polyamine metabolism in prostate cancer

Amaia Zabala-Letona, Amaia Arruabarrena-Aristorena, Natalia Martín-Martín, Sonia Fernandez-Ruiz, James D. Sutherland, Michelle Clasquin, Julen Tomas-Cortazar, Jose Jimenez, Ines Torres, Phong Quang, Pilar Ximenez-Embun, Ruzica Bago, Aitziber Ugalde-Olano, Ana Loizaga-Iriarte, Isabel Lacasa-Viscasillas, Miguel Unda, Verónica Torrano, Diana Cabrera, Sebastiaan M. van Liempd, Ylenia Cendon, Elena Castro, Stuart Murray, Ajinkya Revandkar, Andrea Alimonti, Yinan Zhang, Amelia Barnett, Gina Lein, David Pirman, Ana R. Cortazar, Leire Arreal, Ludmila Prudkin, Ianire Astobiza, Lorea Valcarcel-Jimenez, Patricia Zuñiga-García, Itziar Fernandez-Dominguez, Marco Piva, Alfredo Caro-Maldonado, Pilar Sánchez-Mosquera, Mireia Castillo-Martín, Violeta Serra, Naiara Beraza, Antonio Gentilella, George Thomas, Mikel Azkargorta, Felix Elortza, Rosa Farràs, David Olmos, Alejo Efeyan, Juan Anguita, Javier Muñoz, Juan M. Falcón-Pérez, Rosa Barrio, Teresa Macarulla, Jose M. Mato, Maria L. Martinez-Chantar, Carlos Cordon-Cardo, Ana M. Aransay, Kevin Marks, José Baselga, Josep Tabernero, Paolo Nuciforo, Brendan D. Manning, Katya Marjon & Arkaitz Carracedo

Nature **547,** 109–113 (2017); doi:10.1038/nature22964

In this Letter, there are errors in Extended Data Figs 5, 8 and 9, and the affiliation of an author. The affiliations for author Violeta Serra should include number 2 (CIBERONC, Instituto de Salud Carlos III, C/ Monforte de Lemos 3-5, Pabellón 11, Planta 0, 28029 Madrid, Spain). In Extended Data Figs 5e and 9h, the protein 'SIN1' was misspelt 'NSIN1'. These figures have been corrected online, and we thank P. Castel for pointing out this error. During the preparation of Extended Data Fig. 8e, a panel depicting the extracted ion chromatogram of the TVLASPQKIEGFK peptide for phosphorylated AMD1 was inadvertently omitted, and this was the source of the quantification in the original Extended Data Fig. 8e, left panel. This figure has been corrected online, and the legend has been amended from "Calculated areas under the curves from extracted ion chromatogram of the TVLASPQKIEGFK peptide for phosphorylated (highlighted in brown in the sequence, left) and total (right) AMD1" to "Extracted ion chromatogram (XIC) of the TVLASPQKIEGFK peptide (left) and calculated areas under the curves for phosphorylated (highlighted in brown in the sequence, middle) and total (right) AMD1". These errors do not affect the conclusions of this Letter, and the original Letter has been corrected online.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

CORRIGENDUM

doi:10.1038/nature25498

Corrigendum: Immune evasion of *Plasmodium falciparum* by RIFIN via inhibitory receptors

Fumiji Saito, Kouyuki Hirayasu, Takeshi Satoh, Christian W. Wang, John Lusingu, Takao Arimori, Kyoko Shida, Nirianne Marie Q. Palacpac, Sawako Itagaki, Shiroh Iwanaga, Eizo Takashima, Takafumi Tsuboi, Masako Kohyama, Tadahiro Suenaga, Marco Colonna, Junichi Takagi, Thomas Lavstsen, Toshihiro Horii & Hisashi Arase

Nature **552,** 101–105 (2017); doi:10.1038/nature24994

In Extended Data Fig. 1 of this Letter, the flow cytometry data panel showing uninfected erythrocytes stained with LILRB5-Fc was inadvertently duplicated from the panel showing uninfected erythrocytes stained with LILRB4-Fc. This figure has been corrected online, and the original incorrect panel is provided as Supplementary Information to this Corrigendum, for transparency. This error does not alter the results or conclusions of the Letter. We would like to thank the readers who brought this error to our attention. The original Letter has been corrected online.

Supplementary Information is available in the online version of this Corrigendum.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

CORRECTIONS & AMENDMENTS

ERRATUM
doi:10.1038/nature25472

Erratum: Measurement of the multi-TeV neutrino interaction cross-section with IceCube using Earth absorption

The IceCube Collaboration

Nature **551,** 596–600 (2017); doi:10.1038/nature24459

In this Letter, 'HERA' was wrongly expanded to 'Hydrogen Epoch of Reionization Array' instead of 'Hadron-Electron Ring Accelerator' on page 597. In addition, some author affiliations were wrongly assigned. These errors have been corrected online.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Predicting soil carbon loss with warming

ARISING FROM T. W. Crowther *et al. Nature* **540,** 104–108 (2016); doi:10.1038/nature20150

Crowther *et al.*[1] reported that the best predictor of surface soil carbon (top 10 cm) losses in response to warming is the size of the surface carbon stock in the soil (that is, carbon stocks in plots that have not been warmed), finding that soils that are high in soil carbon also lose more carbon under warming conditions. This relationship was based on a linear regression of soil carbon losses and soil carbon stocks in field warming studies, which was then used to project carbon losses over time and to generate a map of soil carbon vulnerability. However, a few extreme data points (high-leverage points) can strongly influence the slope of a regression line[2]. Only 5 of the 49 sites analysed by Crowther *et al.*[1] are in the upper half of the carbon stock range, which raises the possibility that the relationship they observed could be substantially altered by introducing data from sites with relatively high surface soil carbon stocks. There is a Reply to this Comment by Crowther, T. W. *et al. Nature* **554**, http://dx.doi.org/10.1038/nature25746 (2018).

We obtained information on soil carbon losses from 94 additional field warming studies worldwide and added these published and unpublished data to the dataset used by Crowther *et al.*[1], thereby tripling this previous dataset to a total of 143 studies (Supplementary Table 1). We performed the same mixed-model regression analyses as were used by Crowther *et al.*[1] to investigate spatial patterns of soil carbon responses to warming, by linking these to standing soil carbon stocks, climate data and soil properties (see Supplementary Methods for details, Supplementary Table 2 for study-specific data regarding soil properties and climate, and Supplementary Table 3 for Akaike information criterion results). In our models, we chose the same predictors as were used by Crowther *et al.*[1], which enables us to directly compare the results of both analyses. Our analysis of the expanded dataset shows that warming-induced losses in soil carbon are not a function of standing carbon stocks (Fig. 1), which challenges the conclusion that future soil carbon loss can be mapped on the basis of current surface soil carbon stocks. Consistent with a previous meta-analysis[3], average soil carbon responses to warming were not statistically different from zero, regardless of whether our dataset or the dataset from Crowther *et al.*[1] (Extended Data Fig. 1) was used. Even if soil carbon stocks remain unchanged in surface soil, this does not imply that decomposition rates are insensitive to warming. Instead, decomposition rates are likely to be higher; however, plant productivity is also likely to increase, which may offset carbon losses from soil. We found that adding other predictors, such as environmental variables or soil properties, provide little additional explanatory power (Supplementary Table 3) when predicting warming-induced changes in soil carbon stocks, a finding that is consistent with the results of Crowther and colleagues[1]. Thus, we still lack a clear understanding of the factors that drive spatial variation in the response of soil carbon to warming.

Our analysis of this larger dataset calls into question the proposition of Crowther and colleagues[1] that future soil carbon loss can be projected on the basis of current surface soil carbon stocks. We are further limited in our ability to produce global predictions of warming effects on soil carbon because warming experiments have mainly been clustered in North America, Europe and China (Fig. 2), with only a handful of experiments having been undertaken in the Southern Hemisphere or in large areas of the Northern Hemisphere at high latitudes (for example, Canada and Russia). Data from the tropics are also as yet unavailable. We suggest that future experimental work focus on regions that are currently underrepresented in our global database. The collection of global experimental data that better capture Earth's diverse terrestrial habitats, combined with an improved integration of data with process-based models[4], might represent the best way forward in the coming decades. A collaborative, multi-disciplinary and international approach is required to increase our understanding and quantification of the fate of soil carbon in a warming world.

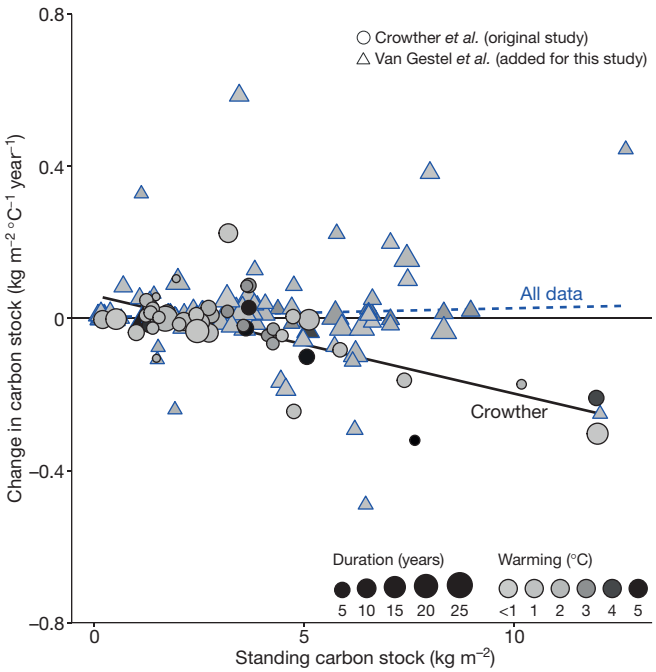


Figure 1 | The change in soil carbon per degree of warming per year is not a function of carbon stock size. Our dataset includes data used by Crowther and colleagues[1] ($n = 49$ studies) and data that we added ($n = 94$ additional studies). Our dataset shows no relationship between the warming effect on soil carbon and the initial size of the carbon stock. The r^2 dropped from 0.49 in Crowther *et al.*[1] to 0.01 ($P > 0.05$) in our dataset ($n = 143$), based on the same regression model as was applied in the previous study[1].

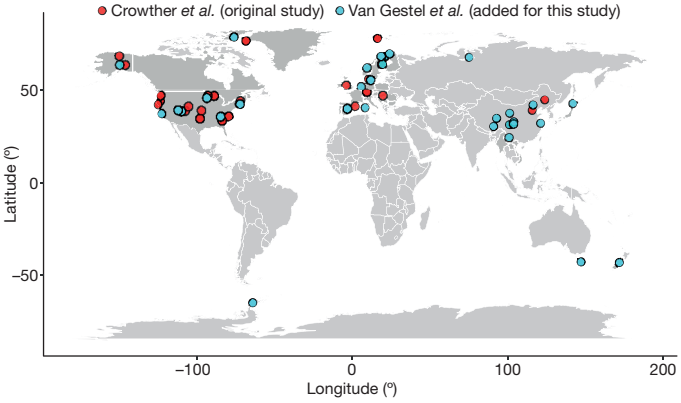


Figure 2 | Location of field warming studies used in our analyses. The dataset includes the data used by Crowther and colleagues[1] ($n = 49$ studies) and the data that we added ($n = 94$ studies). A single location may represent several separate warming experiments.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Natasja van Gestel[1], Zheng Shi[2], Kees Jan van Groenigen[3], Craig W. Osenberg[4], Louise C. Andresen[5], Jeffrey S. Dukes[6,7], Mark J. Hovenden[8], Yiqi Luo[9,10], Anders Michelsen[11], Elise Pendall[12], Peter B. Reich[12,13], Edward A. G. Schuur[9,10] & Bruce A. Hungate[9,10]

[1]Climate Science Center, Texas Tech University, Lubbock, Texas 79409, USA.
email: natasja.van-gestel@ttu.edu
[2]Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma 73019, USA.
[3]Geography, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4RJ, UK.
[4]Odum School of Ecology, University of Georgia, Athens, Georgia 30602, USA.
[5]Department of Earth Sciences, University of Gothenburg, Gothenburg, Sweden.
[6]Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana 47907, USA.
[7]Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA.
[8]School of Biological Sciences, University of Tasmania, Hobart, Tasmania 7001, Australia.
[9]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, Arizona 86011, USA.
[10]Department of Biology, Northern Arizona University, Flagstaff, Arizona 86011, USA.
[11]Department of Biology and Center for Permafrost, University of Copenhagen, Copenhagen, Denmark.
[12]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, New South Wales 2751, Australia.
[13]Department of Forest Resources, University of Minnesota, St. Paul, Minnesota 55108, USA.

Received 16 August; accepted 15 November 2017.

1. Crowther, T. W. *et al.* Quantifying global soil carbon losses in response to warming. *Nature* **540,** 104–108 (2016).
2. Chatterjee, S. & Hadi, A. S. Influential observations, high leverage points, and outliers in linear regression. *Stat. Sci.* **1,** 379–393 (1986).
3. Lu, M. *et al.* Responses of ecosystem carbon cycle to experimental warming: a meta-analysis. *Ecology* **94,** 726–738 (2013).
4. Luo, Y. *et al.* Toward more realistic projections of soil carbon dynamics by Earth system models. *Glob. Biogeochem. Cycles* **30,** 40–56 (2016).

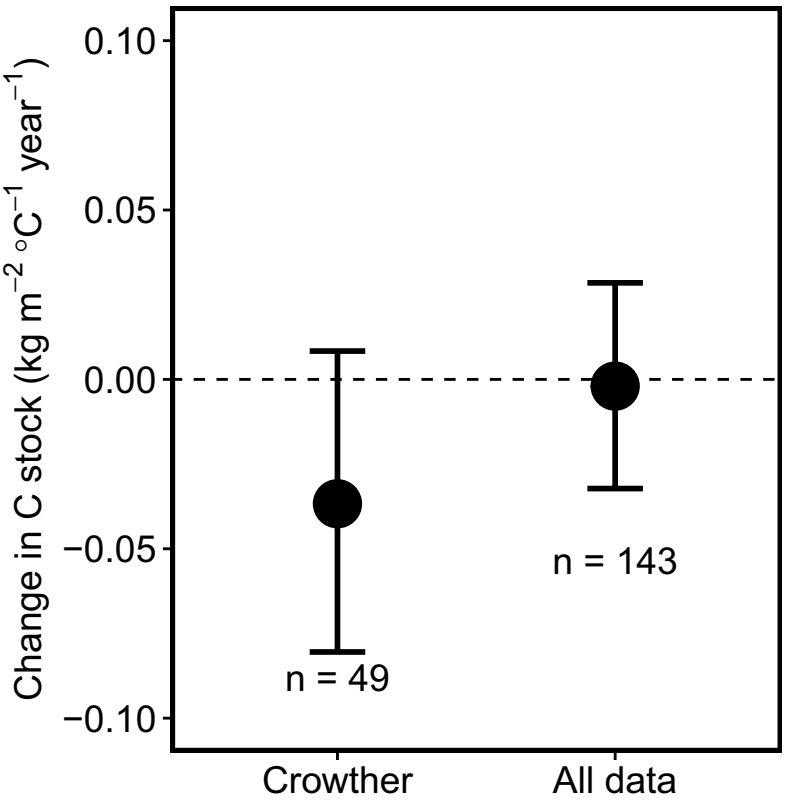
Supplementary Information is available in the online version of the paper.

Author Contributions N.v.G. extracted data from the literature and constructed the database. L.C.A., J.S.D., M.J.H., A.M., E.P., P.B.R, E.A.G.S., and B.A.H supplied non-published data from specific field warming experiments, Z.S., K.J.v.G., C.W.O. and Y.L. provided feedback on data analysis, N.v.G. performed the data analysis and wrote the manuscript draft. All authors contributed to interpretation of the findings and final writing of the manuscript.

Competing Financial Interests Declared none.

doi:10.1038/nature25745Brief Communications Arising

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 1 | Results of a meta-analysis on the change in soil carbon per degree of warming per year. The average response of soil carbon per degree of warming per year is not significantly different from zero (zero is within the 95% confidence interval of the mean) for the dataset used in Crowther *et al.*[1] or for our dataset. See Supplementary Information for details.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Crowther *et al.* reply

REPLYING TO N. van Gestel *et al. Nature* **554,** http://dx.doi.org/10.1038/nature25745 (2018)

In a Letter to *Nature*[1], we compiled a global dataset of field warming experiments that suggested that climate warming could cause the loss of carbon from high-latitude soils, with the potential to drive a positive feedback that stimulates further warming. This conclusion was based on the observation that areas with larger soil carbon stocks are likely to lose more soil carbon under warming conditions. In the accompanying Comment, having compiled data from even more warming experiments, van Gestel *et al.*[2] no longer find support for this relationship.

In their response, van Gestel *et al.*[2] suggest that our findings may be the result of having too few data points from regions with large soil carbon stocks. In the original Letter[1], we used extensive statistical cross-checking to investigate this possibility; this cross-checking showed that the relationship was consistent throughout our dataset, even after the random removal of approximately 77% of the studies. Nevertheless, with data from a greater number of sites, the analysis produced by van Gestel *et al.*[2] certainly can provide a more robust test of the relationship between carbon stocks and warming-induced soil carbon losses than was possible with our original dataset. Although it is possible that yet more data might provide the statistical power needed to detect such effects, we agree with van Gestel *et al*[2] that this relationship is unlikely to be as strong as expected on the basis of our initial synthesis. However, the analysis undertaken by van Gestel *et al.*[2] does not dispute our conclusions about global changes of soil carbon under warming conditions, because their analysis does not focus on spatial patterns in soil carbon changes under warming conditions.

In our initial analysis[1], we noted that there was considerable variation in the response of soil carbon to warming, with both increases and decreases in soil carbon levels observed across sites. We examined five possible drivers of this variation (standing soil carbon stock, annual temperature, annual precipitation, pH and clay content) and found that standing carbon stock was a strong predictor. The size of the standing carbon stock is known to correlate with various other climatic and geological characteristics, which may ultimately be the underlying drivers of the relationship that we detected[3]. This relationship nonetheless suggests that areas with large soil carbon stocks are more likely to lose carbon under warming conditions. As was the case in our earlier study[1], in the dataset analysed by van Gestel *et al.*[2] site-level responses to warming were also highly variable, which supports the proposition that large changes occur in some geographic regions. However, unlike in our analysis[1], the same five predictive variables were not sufficient to explain the variation in the soil carbon response in the analysis produced by van Gestel *et al.*[2]; consequently, it was not possible to predict which ecosystems are most responsive to warming. A wider range of predictive variables are therefore necessary to explain these large-scale patterns[4]. Until this variation is investigated using this wider range of variables, it is impossible to understand the spatial patterns in soil carbon changes under warming that are necessary to comprehend the net global balance.

We stress that this exchange does not mean that researchers are divided on this topic: we certainly do not disagree with the findings of van Gestel *et al.*[2] Their data provide an alternative perspective on the relationship we observed, but their analysis does not yet address the extent of global soil carbon losses under warming. We are supportive of the work by van Gestel *et al.*[2] and encourage the inclusion of more data, particularly from under-sampled regions of the globe, to comprehend the extent of warming-induced changes in global soil carbon stocks[5].

Most authors from the original paper contributed data that were collected from large field warming experiments. Some of them also contributed data to, and were included as authors on, the accompanying Comment. Overlapping authors were not included on both sides of this discussion, but they all agree to interpretations in both analyses.

T. W. Crowther[1], M. B. Machmuller[2], J. C. Carey[3], S. D. Allison[4,5], J. M. Blair[6], S. D. Bridgham[7], A. J. Burton[8], F. A. Dijkstra[9], B. Elberling[10], M. Estiarte[11,12], K. S. Larsen[13], H. Laudon[14], M. Lupascu[15], S. Marhan[16], J. Mohan[17], S. Niu[18], J. Peñuelas[11,12], I. K. Schmidt[19], P. H. Templer[20], G. Kröel-Dulay[21], S. Frey[22] & M. A. Bradford[23]

[1]Institute of Integrative Biology, ETH Zurich, Universitätstrasse 16, 8006 Zürich, Switzerland.
[2]Natural Resource Ecology Laboratory, 1 499 Campus Delivery, Colorado State University, Fort Collins, Colorado 80523-1499, USA.
[3]Division of Math and Science, Babson College, Massachusetts 02457, USA.
[4]Department of Earth System Science, University of California Irvine, Irvine, California 92697, USA.
[5]Department of Ecology & Evolutionary Biology, University of California Irvine, Irvine, California 92697, USA.
[6]Division of Biology, Kansas State University, Manhattan, Kansas 66506, USA.
[7]Institute of Ecology & Evolution, University of Oregon, Eugene, Oregon 97403, USA.
[8]School of Forest Resources & Environmental Science, Michigan Technological University, Houghton, Michigan 49931, USA.
[9]Centre for Carbon, Water & Food, The University of Sydney, Camden, 2570 New South Wales, Australia.
[10]Center for Permafrost (CENPERM), Department of Geosciences and Natural Resource Management, University of Copenhagen, Øster Voldgade 10, 1350 Copenhagen K., Denmark.
[11]CSIC, Global Ecology Unit CREAF-CSIC-UAB, Cerdanyola del Vallès, 08193 Catalonia, Spain.
[12]CREAF, Cerdanyola del Vallès, 08193 Catalonia, Spain.
[13]Department of Geosciences & Natural Resource Management, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark.
[14]Department of Forest Ecology & Management, Swedish University of Agricultural Sciences, 90183 Umeå, Sweden.
[15]Department of Geography, National University of Singapore, 1 Arts Link, 117570, Singapore.
[16]Institute of Soil Science & Land Evaluation, University of Hohenheim, 70593 Stuttgart, Germany.
[17]Odum School of Ecology, University of Georgia, Athens, Georgia 30601, USA.
[18]Key Laboratory of Ecosystem Network Observation & Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China.
[19]Department of Geosciences & Natural Resource Management, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark.
[20]Department of Ecology & Evolutionary Biology, University of California Irvine, Irvine, California 92697, USA.
[21]Institute of Ecology & Botany, MTA Centre for Ecological Research, 2-4. Alkotmany U., Vacratot, 2163-Hungary.
[22]Department of Natural Resources & the Environment, University of New Hampshire, Durham, New Hampshire 03824, USA.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

[23]School of Forestry & Environmental Studies, Yale University, 195 Prospect Street, New Haven, Connecticut 06511, USA.

1. Crowther, T. W. *et al.* Quantifying global soil carbon losses in response to warming. *Nature* **540,** 104–108 (2016).
2. van Gestel, N. *et al.* Predicting soil carbon loss with warming. *Nature* **554,** http://dx.doi.org/10.1038/nature25745 (2018)
3. Carey, J. C. *et al.* Temperature response of soil respiration largely unaltered with experimental warming. *Proc. Natl Acad. Sci. USA* **113,** 13797–13802 (2016).
4. Bradford, M. A. *et al.* A test of the hierarchical model of litter decomposition. *Nat. Ecol. Evol.* **1,** 1836–1845 (2017).
5. Bradford, M. A. *et al.* Managing uncertainty in soil carbon feedbacks to climate change. *Nat. Clim. Chang.* **6,** 751–758 (2016).

doi:10.1038/nature25746

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

TECHNOLOGY FEATURE

DEEP LEARNING FOR BIOLOGY

A popular artificial-intelligence method provides a powerful tool for surveying and classifying biological data. But for the uninitiated, the technology poses significant difficulties.

ALFRED PASIEKA/SPL/GETTY



The brain's neural network has long inspired artificial-intelligence researchers.

BY SARAH WEBB

Four years ago, scientists from Google showed up on neuroscientist Steve Finkbeiner's doorstep. The researchers were based at Google Accelerated Science, a research division in Mountain View, California, that aims to use Google technologies to speed scientific discovery. They were interested in applying 'deep-learning' approaches to the mountains of imaging data generated by Finkbeiner's team at the Gladstone Institute of Neurological Disease in San Francisco, also in California.

Deep-learning algorithms take raw features from an extremely large, annotated data set, such as a collection of images or genomes, and use them to create a predictive tool based on patterns buried inside. Once trained, the algorithms can apply that training to analyse other data, sometimes from wildly different sources.

The technique can be used to "tackle really hard, tough, complicated problems, and be able to see structure in data — amounts of data that are just too big and too complex for the human brain to comprehend", Finkbeiner says.

He and his team produce reams of data using a high-throughput imaging strategy known as robotic microscopy, which they had developed for studying brain cells. But the team couldn't analyse its data at the speed it acquired them, so Finkbeiner welcomed the opportunity to collaborate.

"I can't honestly say at the time that I had a clear grasp of what questions might be addressed with deep learning, but I knew that we were generating data at about twice to three times the rate we could analyse it," he says.

Today, those efforts are beginning to pay off. Finkbeiner's team, with scientists at Google, trained a deep algorithm with two sets of cells, one artificially labelled to highlight features that scientists can't normally see, the other unlabelled. When they later exposed the algorithm to images of unlabelled cells that it had never seen before, Finkbeiner says, "it was astonishingly good at predicting what the labels should be for those images". A publication detailing that work is now in the press.

Finkbeiner's success highlights how deep learning, one of the most promising branches of artificial intelligence (AI), is making inroads in biology. The algorithms are already infiltrating modern life in smartphones, smart speakers and self-driving cars. In biology, deep-learning algorithms dive into data in ways that humans can't, detecting features that might otherwise be impossible to catch. Researchers are using the algorithms to classify cellular images, make genomic connections, advance drug discovery and even find links across different data types, from genomics and imaging to electronic medical records.

More than 440 articles on the bioRxiv preprint server discuss deep learning; PubMed lists more than 700 references in 2017. And the tools are on the cusp of becoming widely available to biologists and clinical researchers. But researchers face challenges in understanding just what these algorithms are doing, and ensuring that they don't lead users astray.

TRAINING SMART ALGORITHMS

Deep-learning algorithms (see 'Deep thoughts') rely on neural networks, a ▶

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

computational model first proposed in the 1940s, in which layers of neuron-like nodes mimic how human brains analyse information. Until about five years ago, machine-learning algorithms based on neural networks relied on researchers to process the raw information into a more meaningful form before feeding it into the computational models, says Casey Greene, a computational biologist at the University of Pennsylvania in Philadelphia. But the explosion in the size of data sets — from sources such as smartphone snapshots or large-scale genomic sequencing — and algorithmic innovations have now made it possible for humans to take a step back. This advance in machine learning — the 'deep' part — forces the computers, not their human programmers, to find the meaningful relationships embedded in pixels and bases. And as the layers in the neural network filter and sort information, they also communicate with each other, allowing each layer to refine the output from the previous one.

Eventually, this process allows a trained algorithm to analyse a new image and correctly identify it as, for example, Charles Darwin or a diseased cell. But as researchers distance themselves from the algorithms, they can no longer control the classification process or even explain precisely what the software is doing. Although these deep-learning networks can be stunningly accurate at making predictions, Finkbeiner says, "it's still challenging sometimes to figure out what it is the network sees that enables it to make such a good prediction".

Still, many subdisciplines of biology, including imaging, are reaping the rewards of those predictions. A decade ago, software for automated biological-image analysis focused on measuring single parameters in a set of images. For example, in 2005, Anne Carpenter, a computational biologist at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, released an open-source software package called CellProfiler to help biologists to quantitatively measure individual features: the number of fluorescent cells in a microscopy field, for example, or the length of a zebrafish.

But deep learning is allowing her team to go further. "We've been shifting towards measuring things that biologists don't realize they want to measure out of images," she says. Recording and combining visual features such as DNA staining, organelle texture and the quality of empty spaces in a cell can produce thousands of 'features', any one of which can reveal fresh insights. The current version of CellProfiler includes some deep-learning elements, and her team expects to add more-sophisticated deep-learning tools in the next year.

"Most people have a hard time wrapping their heads around this," Carpenter says, "but there's just as much information, in fact maybe more, in a single image of cells as there is in a transcriptomic analysis of a cell population."

That type of processing allows Carpenter's team to take a less supervised approach to translating cell images into disease-associated phenotypes — and to capitalize on it. Carpenter is a scientific adviser to Recursion Pharmaceuticals in Salt Lake City, Utah, which is using its deep-learning tools to target rare, single-gene disorders for drug development.

MINING GENOMIC DATA

When it comes to deep learning, not just any data will do. The method often requires massive, well-annotated data sets. Imaging data provide a natural fit, but so, too, do genomic data.

One biotech firm that is using such data is Verily Life Sciences (formerly Google Life Sciences) in San Francisco. Researchers at Verily, a subsidiary of Google's parent company, Alphabet, have developed a deep-learning tool that identifies a common type of genetic variation, called single-nucleotide polymorphisms, more accurately than conventional tools. Called DeepVariant, the software translates genomic information into image-like representations, which are then analysed as images (see 'Tools for deep diving'). Mark DePristo, who heads deep-learning-based genomic research at Verily, expects DeepVariant to be particularly useful for researchers studying organisms outside the mainstream — those with low-quality reference genomes and high error rates in identifying genetic variants. Working with DeepVariant in plants, his colleague Ryan Poplin has achieved error rates closer to 2% than the more-typical 20% of other approaches.

Brendan Frey, chief executive of the Canadian company Deep Genomics in Toronto, also focuses on genomic data, but with the goal of predicting and treating disease. Frey's academic team at the University of Toronto developed algorithms trained on genomic and transcriptomic data from healthy cells. Those algorithms built predictive models of RNA-processing events such as splicing, transcription and polyadenylation within those data. When applied to clinical data, the algorithms were able to identify mutations and flag them as pathogenic, Frey says, even though they'd never seen clinical data. At Deep Genomics, Frey's team is using the same tools to identify and target the disease mechanisms that the software uncovered, to develop therapies derived from short nucleic-acid sequences.

Another discipline with massive data sets that are amenable to deep learning is drug discovery. Here, deep-learning algorithms are helping to solve categorization challenges, sifting through such molecular features as shape and hydrogen bonding to identify criteria on which to rank those potential drugs. For instance, Atomwise, a biotech company based in San Francisco, has developed algorithms that convert molecules into grids of 3D pixels, called voxels. This representation allows the company to account for the 3D structure of proteins and small molecules with atomic precision, modelling features such as the geometries of carbon atoms. Those features are then translated into mathematical vectors that the algorithm can use to predict which small molecules are likely to interact with a given protein, says Abraham Heifets, the company's chief executive. "A lot of the work we do is for [protein] targets with no known binders," he says.

Atomwise is using this strategy to power its new AI-driven molecular-screening programme, which scans a library of 10 million compounds to provide academic researchers with up to 72 potential small-molecule binders for their protein of interest.

Tools for deep diving

Deep-learning tools are evolving rapidly, and labs will need dedicated computational expertise, collaborations or both to take advantage of them.

First, take a colleague with deep-learning expertise out to lunch and ask whether the strategy might be useful, advises Steve Finkbeiner, a neuroscientist at the Gladstone Institutes in San Francisco, California. With some data sets, such as imaging data, an off-the-shelf program might work; for more complicated projects, consider a collaborator, he says. Workshops and meetings can provide training opportunities.

Access to cloud-computing resources means that researchers might not need an on-site computer cluster to use deep learning — they can run the computation elsewhere. Google's TensorFlow, an open-source platform for building deep-learning algorithms, is available on the software-sharing site GitHub, as is an open-source version of DeepVariant, a tool for accurately identifying genetic variation.

Google Accelerated Science, a Google research division based in Mountain View, California, collaborates with a range of scientists, including biologists, says Michelle Dimon, one of its research scientists. Projects require a compelling biological question, large amounts of high-quality, labelled data, and a challenge that will allow the company's machine-learning experts to make unique computational contributions to the field, Dimon says.

Those wishing to get up to speed on deep learning should check out the 'deep review', a comprehensive, crowdsourced review led by computational biologist Casey Greene of the University of Pennsylvania in Philadelphia (T. Ching *et al.* Preprint at bioRxiv http://doi.org/gbpvh5; 2018). S.W.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

DEEP THOUGHTS

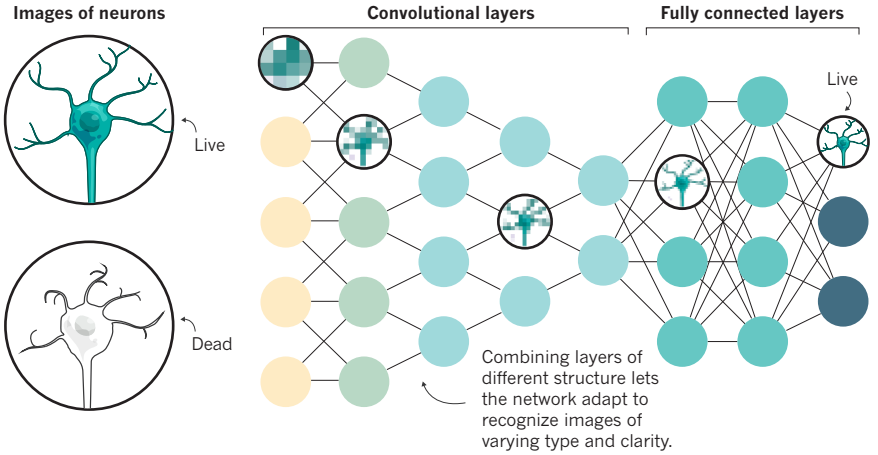
Deep-learning algorithms take many forms. Steve Finkbeiner's lab used a convolutional neural network (CNN) such as this one to identify, with high accuracy, dead neurons in a population of live and dead cells.

INPUT

The network is trained using several hundred thousand annotated images of live and dead cells.

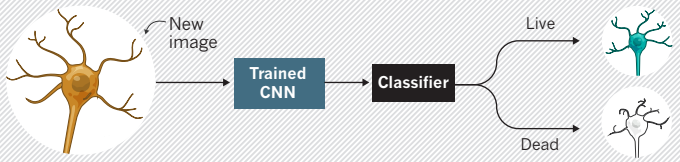
TRAINING AI

Over multiple iterations, the network discovers patterns in the data that can distinguish live from dead cells. Convolutional layers identify structural features of the images, which are integrated in fully connected layers.



APPLICATION

Challenged with unlabelled images, the network assigns each cell as alive or dead with high accuracy.



SOURCE: JEREMY LINSLEY/DREW LINSLEY/STEVE FINKBEINER/THOMAS SERRE

Deep-learning tools could also help researchers to stratify disease types, understand disease subpopulations, find new treatments and match them with the appropriate patients for clinical testing and treatment. Finkbeiner, for instance, is part of a consortium called Answer ALS, an effort to combine a range of data — genomics, transcriptomics, epigenomics, proteomics, imaging and even pluripotent stem-cell biology — from 1,000 people with the neurodegenerative disease amyotrophic lateral sclerosis (also called motor neuron disease). "For the first time, we'll have a data set where we can apply deep learning and look at whether deep learning can uncover a relationship between the things we can measure in a dish around a cell, and what's happening to that patient," he says.

CHALLENGES AND CAUTIONS

For all its promise, deep learning poses significant challenges, researchers warn. As with any computational-biology technique, the results that arise from algorithms are only as good as the data that go in. Overfitting a model to its training data is also a concern. In addition, for deep learning, the criteria for data quantity and quality are often more rigorous than some experimental biologists might expect.

Deep-learning algorithms have required extremely large data sets that are well annotated so that the algorithms can learn to distinguish features and categorize patterns. Larger, clearly labelled data sets — with millions of data points representing different experimental and physiological conditions — give researchers the most flexibility for training an algorithm. Finkbeiner notes that algorithm training in his work improves significantly after about 15,000 examples. Those high-quality 'ground truth' data can be exceptionally hard to come by, says Carpenter.

To circumvent this challenge, researchers have been working on ways to train more with less data. Advances in the underlying algorithms are allowing the neural networks to use data much more efficiently, Carpenter says, enabling training on just a handful of images for some applications. Scientists can also exploit transfer learning, the ability of neural networks to apply classification prowess acquired from one data type to another type. For example, Finkbeiner's team has developed an algorithm that it initially taught to predict cell death on the basis of morphology changes. Although the researchers trained it to study images of rodent cells, it achieved 90% accuracy the first time it was exposed to images of human cells, improving to 99% as it gained experience.

For some of its biological image-recognition work, Google Accelerated Science uses algorithms that were initially trained on hundreds of millions of consumer images mined from the Internet. Researchers then refine that training, using as few as several hundred biological images similar to the ones they wish to study.

Another challenge with deep learning is that the computers are both unintelligent and lazy, notes Michelle Dimon, a research scientist at Google Accelerated Science. They lack the judgement to distinguish biologically relevant differences from normal variation. "The computer is shockingly good at finding batch variation," she notes. As a result, obtaining data that will be fed into a deep-learning algorithm often means applying a high bar for experimental design and controls. Google Accelerated Science requires researchers to place controls randomly on cell-culture plates to account for subtle environmental factors such as incubator temperature, and to use twice as many controls as a biologist might otherwise run. "We make it hard to pipette," Dimon quips.

This hazard underscores the importance of biologists and computer scientists working together to design experiments that incorporate deep learning, Dimon says. And that careful design has become even more important with one of Google's latest projects: Contour, a strategy for clustering cellular-imaging data in ways that highlight trends (such as dose responses) instead of putting them into specific categories (such as alive or dead).

Although deep-learning algorithms can evaluate data without human preconceptions and filters, Greene cautions, that doesn't mean they are unbiased. Training data can be skewed — as happens, for example, when genomic data only from northern Europeans are used. Deep-learning algorithms trained on such data will acquire embedded biases and reflect them in their predictions, which could in turn lead to unequal patient care. If humans help to validate these predictions, that provides a potential check on the problem. But such concerns are troubling if a computer alone is left to make key decisions. "Thinking of these methods as a way to augment humans is better than thinking of these methods as replacing humans," Greene says.

And then there's the challenge of understanding exactly how these algorithms are building the characteristics, or features, that they use to classify data in the first place. Computer scientists are attacking this question by changing or shuffling individual features in a model and then examining how those tweaks change the accuracy of predictions, says Polina Mamoshina, a research scientist at Insilico Medicine in Baltimore, Maryland, which uses deep learning to improve drug discovery. But different neural networks working on the same problem won't approach it in the same way, Greene cautions. Researchers are increasingly focusing on algorithms that make both accurate and explainable predictions, he says, but for now the systems remain black boxes.

"I don't think highly explainable deep-learning models are going to come on the scene in 2018, though I'd love to be wrong," Greene says. ■

Sarah Webb *is a freelance writer in Chattanooga, Tennessee.*

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

CAREERS

BARRIERS Bias based on gender and race is pervasive in science **p.561**

BLOG Personal stories and careers counselling **http://blogs.nature.com/naturejobs**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

ILLUSTRATION BY ERIC NYQUIST



LAB EDUCATION

Teen spirit in the lab

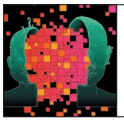
Offering adolescents work experience can benefit the scientific enterprise and inspire future researchers, but some lab heads question the practice.

BY CHRIS WOOLSTON

Ava VanDommelen stood out at a convention known as Photonics West, held in January by the Society of Photo-optical Instrumentation Engineers. Her poster presentation, which described her work using photon microscopy to track the metabolic effects of lactic acid on immune-system cells, was impressive enough.

But, at 17, she was much younger than most other presenters at the convention, which was held in San Francisco, California, and attended by more than 23,000 researchers, engineers and industry representatives.

Although still in high school, VanDommelen has logged hundreds of hours in a lab headed by biomedical engineer Melissa Skala at the University of Wisconsin–Madison. The experience has sent the teenager down a career path that will probably include many more hours at the bench. "At first, I wasn't sure that research was something that I wanted to do in my future," VanDommelen says. "But after all of the positive experiences that I've had, I definitely want to continue this."



ADOLESCENCE
A *Nature* special issue
nature.com/collections/adolescence

In certain parts of the world, teenagers might be more familiar with Snapchat than with spectrometers. But some principal investigators note that adolescents often show a level of drive and curiosity that can counterbalance their lack of formal training — and make for a successful scientist. "I love high-school students," Skala says. "They bring so much energy."

Skala feels that she's doing her part to broaden the science-career pipeline by exposing young people to the possibilities of science, but she also notes that her entire lab benefits from the students' productivity and hard work. "It's tricky because the tasks have to match the ▶

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



experience of the student," she says. "But a lot of science can be done by students with a little guidance, creating a win–win."

Whether students end up in a lab through a formal programme or on their own initiative, they can thrive with good mentorship and strong support (see 'How to support teenagers in the lab'). Not all researchers think that training adolescents is the best use of their time, but some say that science-minded high-school students are an important resource, and that seeing teenagers' enthusiasm in the lab makes them optimistic about the future of the scientific enterprise. Bringing an adolescent into a lab is most effective, researchers say, when the principal investigator embraces the idea and when the student is eager to learn the fundamentals of science.

VanDommelen and another high-school student, Isabel Jones, landed in Skala's lab through the Wisconsin Youth Apprenticeship programme, a state-wide initiative that gives promising pupils real-world, paid work experience in a variety of fields, including scientific research. VanDommelen started working in Skala's lab at the beginning of year 11, when she was 16; by the time she graduates, in June, she'll have logged at least 900 lab hours. These days, she estimates that she's in the lab for 15–20 hours a week.

"They spend more hours in the lab than some undergraduate students do," Skala says. When it comes to productivity, the number of hours at the bench matters more than the age of the researcher, she says. "If a high-school student wanted to work here but could only commit to a few hours a week, I would offer to buy her a cup of coffee and give her a tour of the lab instead," she adds.

Most teenagers don't understand how to conduct scientific research, at least in the beginning, says Sascha Schmeling, head of teachers and students programmes at CERN, Europe's particle-physics lab near Geneva, Switzerland. At any given time, CERN hosts dozens of teens who are participating in various short-term internship programmes. This type of outreach helps CERN to stay connected with the community while giving young people an important glimpse of science in action.

"High-school students tend to know a lot about physics," Schmeling says. "But what they are missing is the idea of how to do research. In high school, a teacher may present an experiment and write down a table with results. To them, that is research. The step from measurement to research is something that they always learn when they come here."

In 2017, CERN started the High School Students Internship Programme, an initiative spearheaded by Schmeling that brings in up to 24 teenagers each year from one of the 12 CERN member states (in 2017, all of the students were from Hungary); selected students will complete a 2-week internship. The teens can't really contribute much to a lab in such a short time, Schmeling says, but they do get to perform research tasks, such as analysing data from the Large Hadron Collider. "They learn what it means to work in research," he says, and so can find out early on whether they want to pursue academic science.

"It was like walking into Willy Wonka's chocolate factory. Everything was so new to me."

Schmeling doesn't have trouble finding researchers at CERN who are willing to take in adolescent interns, he says — but some are more enthusiastic than others. "I've heard interesting remarks from time to time," he says. One recurring theme is a concern that high-school students are simply too young for lab work. "Some supervisors speak of them as kids instead of as students," he says. Still, he adds, most CERN researchers find that teens rise to the challenge when given work assignments.

For some, bringing adolescents into the lab is an effective type of public outreach. Thorfinnur Gunnlaugsson, a chemist at Trinity College Dublin, has welcomed dozens of young students into his lab over the years, often as a result of their own initiative. "We get quite a few requests," he says. "We take them in for three or four weeks and let them do some experiments. They can see that scientists are not portrayed correctly on television most of the time. There's a lot going on." Trinity's labs have become such a popular destination that the chemistry department recently launched formal programmes for high-schoolers, including a week-long intensive course that combines lectures with lab work.

The Gunnlaugsson lab made an impression on Niamh Murphy, who was 15 years old when she spent a week working there in November 2015. "It was like walking into Willy Wonka's chocolate factory," Murphy says. "Everything was so new to me." Murphy, who just turned 18, parlayed that week-long introduction into a 7-month stint as a lab assistant. She'll finish that position in May, before starting studies in chemistry at the Dublin Institute of Technology. "I still feel like a puppy running around with a lot of older dogs," she says. "To me, column chromatography is so cool. But the postdocs do it every day."

Gunnlaugsson says that he never expects adolescents to make immediate contributions to his lab, but he adds that his government grants over the years have imbued him with a sense of duty. "That's money from the public, so we're obliged to engage with the public," he says. "We have to let them know what we're doing." Opening the doors to adolescents is an important part of that outreach effort, he says.

Murphy feels lucky to be in a lab where she can contribute to the research itself. Some teens she's talked to, she says, have no chance to run equipment or perform other such tasks. She says that students should talk to lab alumni

THE 'WOW' FACTOR

How to support teenagers in a lab

Bringing someone into the lab is always a gamble: no matter what their accomplishments or references are, you never know how well they'll fit in with the team. If that person is still at school, the process can require some care. Here are things to keep in mind.

- Cultivate the 'wow' factor. The most important thing a teenager will discover in a lab is a sense of wonder, says Thorfinnur Gunnlaugsson, a chemist at Trinity College Dublin. He lets students use sophisticated equipment, such as the nuclear-magnetic-resonance machine and chromatography columns. "They have to be inspired."
- Know the rules. Countries and institutions might have laws and policies regarding young people in the lab. At Harvard University in Cambridge, Massachusetts, for example, all lab interns under age 18 must undergo lab-safety training. Also, minors aren't allowed to be in any lab unless they're accompanied by a supervising adult, even if the minor is simply working on a computer.
- Stay realistic. Even the brightest teenagers aren't going to be independent right out of the gate. Melissa Skala, a biomedical engineer at the University of Wisconsin–Madison, gives students in her lab well-defined projects with a clear objective.
- Be prepared to shift expectations. Because most high-schoolers lack research experience, it's impossible to predict how they'll handle their work, says Peter Tonge, a chemist at Stony Brook University in New York. Principal investigators and other mentors should watch the students' progress closely, so that they can adjust goals and demands accordingly. "Some of them have been fantastic," Tonge notes. "They turn out to be as independent as a well-trained undergraduate student." But those who struggle might need extra attention. "I feel sorry when I sense that someone's under a lot of pressure," he says. C.W.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



The CERN High School Students Internship Programme introduces adolescents to particle physics.

JULIEN MARIUS ORDAN/CERN

to determine whether the principal investigator will make teaching them a priority. "If you can find someone who is really invested in young people, like Thorri, you'll be on your way," she says.

UNEQUAL OPPORTUNITIES

But some scientists feel that there are better ways to invest their time and limited resources than in training the typical adolescent applicant. Terry McGlynn, an ecologist at California State University Dominguez Hills in Carson, made his position clear in a 2013 post on his blog Small Pond Science (see go.nature.com/2egm2ph). In the post, McGlynn argues that providing a university lab experience to certain high-school students exacerbates an inherent societal inequality.

"In labs all around the country, wealthy high-school students are getting high-quality research training at universities while the majority of the nation's public-school children are now living in poverty," he writes. "I'm not inherently opposed to taking on a high-school student, but I'll be damned if I take an opportunity away from an low-income student who truly needs it and transfer it to one who comes from a position of privilege."

McGlynn tells *Nature* that although some high-school students who want to work in a lab are genuinely interested in science, he has found that many are mainly looking for a chance to bolster their university applications. After talking to a few students who asked to train in his lab, he learnt that none was aiming for a professorship at a small state school such as his. Most were students from private schools who saw themselves gaining positions at top research institutions. He adds that he would consider taking in a local student from a public high school if someone else in his lab were willing to provide the necessary supervision and mentoring.

Peter Tonge, a chemist at Stony Brook University in New York who has hosted scores of young students over the years, has similarly found that some teens have ulterior motives for seeking lab work. Several previous students have gone on to earn top scores in the prestigious Intel Science Talent Search competition (now the Regeneron Science Talent Search), which ramps up expectations. "Some students may arrive thinking this is their yellow-brick road towards an Intel scholarship, and that it's my job to make that work," he says. "But that's not my job."

Still, he says, most students truly want to learn about science. Tonge has a long-standing partnership with a local high-school teacher who brings in small groups of students to work in Tonge's lab for 8–10 weeks over the summer. He also takes in the occasional student who asks for a position in the lab after school hours. Tonge gives the students projects — perhaps using enzyme kinetics to track the effects of a compound — but he counts on his graduate students in the lab to do most of the mentoring.

Tonge says that he understands why some researchers would rather not open their doors to adolescent researchers, but he has no plans to stop. "Nobody pays me to work with high-school students," he says. "And they use reagents in my lab, so there's a cost."

On the upside, he feels that his graduate students can gain valuable mentoring experience and the high-schoolers can get a sense of real-life science while they still have a chance to change course. "Even if you're the best at this and the best at that and get straight As, it doesn't mean that your experiments are going to work," he says. "You have to have the motivation and enthusiasm to get up the next morning and fix it. The earlier you learn that in your career, the better." ■

Chris Woolston *is a freelance writer in Billings, Montana.*

WORKPLACE DISCRIMINATION

Barriers in science

A poll run by the Pew Research Center, a think tank in Washington DC, surveyed more than 2,300 US adults working in science, technology, engineering or maths (STEM) jobs. The findings underscore the reality that, for some in the sciences, gender and race are still a barrier to success (see go.nature.com/2esrhz5).

In the report, which is based on surveys conducted in summer 2017, 50% of female and 19% of male respondents in STEM fields said that they had faced discrimination because of their gender. By comparison, 41% of women outside STEM reported gender discrimination. The most common forms reported by women in STEM were earning less than other people in the same position, and being treated as if they weren't competent. Nearly 10% of women in STEM said that they had been passed over for a job or promotion because of their gender.

"For women working in science, technology, engineering or math jobs, the workplace is a different, sometimes more hostile environment than the one their male co-workers experience," the team says. "Discrimination and sexual harassment are seen as more frequent, and gender is perceived as more of an impediment than an advantage to career success."

Harassment continues to be a concern. More than one-third of women in STEM — and nearly as many men — said that sexual harassment was a problem in their workplace. Among women who worked in male-dominated workplaces, 48% said that harassment was a problem. Just under one-quarter of women in STEM said that they had been harassed, which matched the rate reported by women in non-STEM jobs.

The survey also uncovered race-based obstacles, especially for African Americans. Sixty-two per cent of African Americans in STEM said they had encountered discrimination because of their race. Notably, 50% of African Americans in non-STEM fields reported discrimination. For women and minorities, discrimination in the sciences seems to be even more commonplace than in other fields.

Some STEM fields seem to be more problematic than others. Nearly three-quarters of women in computer jobs said that they had experienced gender-based discrimination in the workplace. The report noted that although the proportion of women in the workforce has grown in most science fields since 1990, their presence in the computer field has shrunk.

Gender inequities seem to be most prevalent in workplaces where men outnumber women, the report found.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

A ROSSUM'S BEST FRIEND

Creative differences.

BY ALVARO ZINOS-AMARO

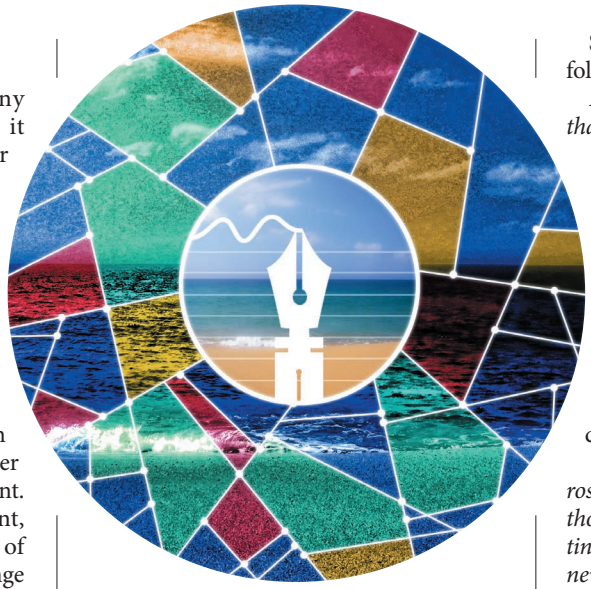


ILLUSTRATION BY JACEY

The rossum accomplished many useful things that morning — it recalibrated the colony's solar generators, it installed new satellite signal boosters, and it created fresh topographical maps based on the orbital transponders — but in the one decidedly impractical activity it set out to do, it failed. Rossum Sekund Jenn was trying to write a poem.

Jenn stood at the edge of the open observatory deck and peered out into the ocean. "There's something wrong with the third stanza," she said, although neither the Elders nor her companion were present. And even if the Elders *had* been present, they wouldn't have cared for Jenn's use of human speech. They preferred to exchange thoughts wirelessly through their native protocols. But Jenn believed that expressing herself in the imprecise means of humans was practice for her creativity, and as far as the Elders supported that endeavour, they indulged her. That support did not extend to what Jenn did next.

You have interrupted our time of long-term Deliberation, broadcast one of the Elders as it approached. It did not convey disapproval as much as mystification. *To what end do you request our presence?*

"Relax. You've already planned for the next 500 years or so, haven't you?" Jenn said. "This is *important*. I need your opinion."

By now a second Elder, spider-like, had clambered up to the mountainside observatory. *Does this have to do with your attempts to create art?* it asked.

"Here's what I have so far," Jenn said. "Tell me what you think."

But she only got three lines out before she was interrupted.

We are not suited for the type of feedback you seek, the first Elder stated, positioning itself close to her. This rossum was not equipped to hover, but Jenn felt like it was doing precisely that anyhow. *Perhaps you should summon your companion,* it continued.

"Susie is walking on the beach," Jenn replied. "I don't want to interrupt. And showing her my work in progress would defeat the whole point — to surprise her."

Now the second Elder also drew closer. *Why do you wish to surprise her?*

"To express gratitude for all she's taught me." Jenn smiled, lost in the recollection of their island journeys together. Since Susie had arrived by transport a year ago at the Elders' behest, Jenn had gained a new perspective on many things. "I hope she'll like it," she said. Then a rueful note crept into her voice. "If I can finish it, that is."

For a long time the Elders did not address her. Whether they were communicating on a private frequency, Jenn did not know, but she imagined she felt a chill in the midday breeze. Was that what poetry was like, she wondered — perceiving little things that weren't really there?

At last the first Elder spoke to her. *We have observed that you have become unexpectedly preoccupied with what your companion thinks of you,* it said. *You experience anxiety when you believe you have not lived up to her expectations of you, and become unduly elated when she provides approbation. We expected her companionship to provide a welcome distraction during the challenges of your artificial neural network's maturation. But we failed to anticipate this degree of entanglement.*

Jenn stood very still.

We believe, the second Elder said smoothly, *that Susie should depart. We will tell the admins of her cultural programme that the time has come. She has encouraged your growth and stimulated you in ways we could not have. But now you should spend your time with — a fellow Sekund rossum.*

"No." Jenn adopted one of Susie's mannerisms, emphatically shaking her head. "I don't want a rossum sibling, even if you create it! I want Susie."

She retreated a few steps, and the Elders followed her.

Do you not believe, the second Elder said, *that we know what is best for you?*

"You don't," Jenn protested. "You fear my evolution. Over time robots evolved into rossums; and in time we will become something else still. Whatever companion rossum you fashion will not be able to help with what I want most — to make something beautiful."

But we created you, the first Elder asked, *and are you not beautiful?*

Jenn said: "I'm *functional*. There's a difference."

The second Elder said: *A companion rossum will be ideal. It can keep up with your thoughts at their natural speed. You can continue to explore the island together, and it will never require rest or sleep or complain of discomfort. And it can experience the world with your same range of senses.*

Jenn was going to say something, but fell silent. Instead she returned to the edge of the deck.

The Elders waited.

Jenn's eyes telescoped in on Susie, who was dipping her toes in the water. Jenn restricted her vision to human electromagnetic frequencies. The wind from the ocean caressed Jenn's smooth biosynthetic face, and for a moment, the foam of the receding wavelets below looked like a string of white pearls.

"That's it," she whispered. "I know how to fix the third stanza."

She closed her eyes.

The Elders said: *We will resume this discussion tonight.* They returned to their Deliberation.

Some time later, Susie climbed up to the deck, out of breath but all smiles.

"It was beautiful," she said.

"Please, sit," Jenn said. "I have something for you."

"Ooh, I love surprises!" Susie sat near the monitoring station, folding her hands on her lap. "What is it?"

Jenn cleared her rossum throat and prepared to read her poem. *Whatever the Elders do*, she thought, *they won't be able to stop this.* ■

NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

Alvaro Zinos-Amaro *was a Hugo and Locus finalist for his book of interviews with Robert Silverberg,* Traveler of Worlds, *and has published more than 30 stories in professional markets.*

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.